Fifth Edition

Essential Surgical Practice

Higher Surgical Training in General Surgery

Edited by
Alfred Cuschieri
George B Hanna

VITALSOURCE® EBOOK

CRC Press
Taylor & Francis Group

Fifth Edition

Essential Surgical Practice

Higher Surgical Training in General Surgery

Fifth Edition

Essential Surgical Practice

Higher Surgical Training in General Surgery

Edited by

Professor Sir Alfred Cuschieri

FRSE, MD, DSc, ChM, FRCS, FACS (Hon), LLD (Hon), FMedSci, FiBiol, Senior Scientist,
Institute for Medical Science and Technology, University of Dundee, Dundee, UK
Professor of Surgery, Scuola Superiore Sant'Anna, Pisa, Italy

Professor George B Hanna

MD, PhD, FRCS, Head of the Division of Surgery,
Imperial College London, UK.

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

CRC Press is an imprint of Taylor & Francis Group, an Informa business

Version Date: 20141031

International Standard Book Number-13: 978-1-4441-3763-7 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

The Fifth Edition of *Essential Surgical Practice* is dedicated to a former Co-Editor of ESP, the late A. R. Moossa, MD (1939–2013), a truly great master surgeon and teacher, and a very close friend who will always be remembered by his Residents brought up within his unforgettable Residency Program at UCSD, California.

Alfred Cuschieri

Contents

PART I: GENERAL SECTION

PART II: SPECIALIST GENERAL SURGERY

Preface

This edition is different from the previous ones in many respects. In the first instance I have recruited a new co-editor, Professor George Hanna, a long-term colleague and friend who is deeply committed to surgical training and education and shares my views on the extreme importance of passing on one's experience and practical clinical knowledge of surgical practice gained over many years to the next generation of surgeons. Between us we have decided to change the structure and content of *Essential Surgical Practice* directing it exclusively to trainees in higher surgical education in General Surgery. This decision reflects the changed scenario in higher surgical training over the past three decades, with the inevitable progress towards specialization in the various Surgical Specialties. Thus all the Specialty Sections included in previous editions (Neurosurgery Thoracic, Vascular and Urology) have been omitted as these are now distinctly separate from General Surgery, which by consensus includes the rest together with Trauma including Head Injuries.

This strategy has allowed us to provide a more detailed account of the various disorders and has enabled us to cover important surgical topics which were either not included in previous editions or were not covered in sufficient depth. We have for this reason recruited new authors, all of whom were carefully selected for their eminence in their respective fields. The 5th edition of *Essential Surgical Practice* is an almost complete rewrite in view of the considerable progress that has taken place in the past 10 years with respect to both pathology and management of the various surgical disorders. The same basic format of *Essential Surgical Practice* has been, however, retained, including highlights of core information for revision purposes and quick reference.

The 5th edition is a single volume with two sections: General section, Chapters 1–15, and Specialist general surgery, Chapters 17–31. *Essential Surgical Practice* will be available in two versions: hard copy and electronic. As with previous editions, both editors have between them written or updated a substantial part of the volume and have taken particular note of all the important recent advances relevant to the subject matter. We have however recruited authors as acknowledged experts for chapters covering topics beyond our clinical expertise, all of whom have provided excellent contributions which required virtually no editing by us.

We hope that the 5th edition of *Essential Surgical Practice* will be well received by the surgical tutors and chiefs of training programmes, and, what is more important to us, that the surgical trainees themselves, irrespective of country, find it a truly useful aid for their surgical education as this has always been our prime objective. Surgical trainees and young consultants are the torch bearers of our profession. In the final analysis, they must always remember that surgery is a craft specialty and that a good surgeon is one who attains fully operating proficiency – nothing less suffices – and, moreover, he or she has the duty of care to his or her patients since technical proficiency and accomplishment are not enough. In essence surgeons are the treatment but the outcome of patients is linked as much to the quality of execution of the operations as to the supportive care which surgeons provide before, during and after surgery.

Finally, we would like to thank our publishers for their assistance, support and especially for their forbearance in extending the deadline for completion a few times. We have no valid excuse for this other than our heavy commitments.

Alfred Cuschieri
George B. Hanna

List of Contributors

Ahmed R. Ahmed BSc(Hons) FRCS (Gen)
Senior Lecturer in Bariatric Surgery, Imperial College London; Consultant Bariatric Surgeon, Clinical Lead - Bariatric Surgery, St Mary's Hospital, Imperial College Healthcare NHS Trust, Academic Division of Surgery, Department of Surgery and Cancer, London

Saba P. Balasubramanian MS PhD FRCS
Clinical Senior Lecturer and Honorary Consultant Endocrine Surgeon, University of Sheffield and Sheffield Teaching Hospitals NHS Foundation Trust

Jane M. Blazeby BSc MSc MD MBChB FRCS(Gen Surg)
Professor of Surgery, University of Bristol and Honorary Consultant Surgeon, University Hospitals Bristol, NHS Foundation Trust, Bristol

Ken Campbell MBChB FRCS(Ed) MD
Consultant Colorectal Surgeon, Ninewells Hospital, Dundee

Arup Chakraborty BSc MBBS MRCP EDIC
Registrar in Anaesthesia and Intensive Care, Princess Alexandra Hospital, Harlow

Raul Coimbra MD PhD FACS
The Monroe E. Trout Professor of Surgery, Executive Vice-Chairman, Department of Surgery; Chief, Division of Trauma, Surgical Critical Care, and Burns, University of California San Diego School of Medicine, San Diego

Richard W. Corbett
Academic Clinical Fellow, Imperial College Renal and Transplant Centre, Hammersmith Hospital, London

John A. Dewar
Consultant and Emeritus Professor of Clinical Oncology at Dundee University and formerly Head of Oncology at Ninewells Hospital, Dundee

John F. Dillon MBBS MD FRCP
Clinical Senior Lecturer, Gut Group, Medical Research Institute, University of Dundee, Ninewells Hospital and Medical School, Dundee

Neill Duncan MBBS MRCP
Renal Consultant, Imperial College Renal and Transplant Centre, Hammersmith Hospital, London

Sam Eljamel MBBCh MD FRCSEd FRCSIre FRCS(SN)
Consultant Neurosurgeon, Ninewells Hospital and Medical School, Dundee

Andrew Evans
Professor of Breast Imaging, Dundee Cancer Centre, University of Dundee, Dundee

Omar Faiz BSc(Hons) MBBS FRCS(GenSurg) MS
Senior Lecturer and Honorary Consultant Surgeon, St Mark's Hospital, Northwick Park, Harro and Academic Surgical Unit, Imperial College Healthcare NHS Trust, London

Edward Fitzgerald BA BM BCh MRCS(Eng)
Registrar in General Surgery, Royal Marsden Hospital, London

Jacob George MD MRCP
Division of Cardiovascular and Diabetes Medicine, Medical Research Institute, University of Dundee Medical School, Dundee

Priya George MBChB MRCOpth MRCP
Division of Cardiovascular and Diabetes Medicine, Medical Research Institute, University of Dundee Medical School, Dundee

Panayiotis Georgiou
NIHR Academic Clinical Fellow in General Surgery, West Middlesex University Hospital and Imperial College Healthcare Trust, London

Mohamad S. Hamady MD FRCR
Consultant Interventional Radiologist and Honorary Senior Lecturer, Department of Interventional Radiology, Imperial College Healthcare NHS Trust; and Department of Surgery and Cancer, Imperial College London

David Hoyt MD FACS
Executive Director, American College of Surgeons, Emeritus Professor of Surgery University of California San Diego School of Medicine, San Diego

Helgi Johannsson BMedSci MBBS MRCP FRCA
Consultant in Anaesthesia and Chief of Service - Theatres and Anaesthesia, St Mary's Hospital, Imperial College Healthcare Trust, London

Leslie Kobayashi MD FACS
Assistant Professor of Surgery, Division of Trauma, Surgical Critical Care, and Burns University of California San Diego Health Sciences, San Diego

Sacha Koch MD PhD
Consultant Surgeon, Ninewells Hospital and Medical School, Dundee

Onn Min Kon MD FRCP
Consultant Respiratory Physician and Honorary Senior Lecturer, Chest and Allergy Clinic, Imperial College School of Medicine, St Mary's Hospital, London

Paul Maddox BSc MCh FRCS(Ed) FRCS(Eng)
Consultant Surgeon, Department of General Surgery, Royal United Hospital, Bath

Vafa Mansoubi MBBS BSc (Hon) FRCA
Specialist Registrar, School of Anaesthesia, Charing Cross Hospital, Imperial College Healthcare Trust

Abdool Rahim Moossa MD FRCS(Eng) FACS (deceased)

Stacy I. Moulder-Thompson
Assistant Professor for Medical Oncology, Department of Breast Medical Oncology, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

William L. Oldfield MSc PhD FRCP
Honorary Clinical Senior Lecturer, National Heart and Lung Institute and Consultant in Respiratory Medicine at Imperial College Healthcare NHS Trust, London

Francesco M. Polignano MD PhD
Unit of HPB and Advanced Laparoscopic Surgery, Department of Surgery and Molecular Oncology, Ninewells Hospital and Medical School, University of Dundee, Dundee

Nicolas Price BMedSci BMBS MRCP FRCA
Speciality Registrar Anaesthesia and Intensive Care, Imperial College NHS Healthcare Trust, St Mary's Hospital, London

Aaron J. Quyn
Clinical Lecturer in Surgery, Clinical and Population Sciences and Education, Ninewells Hospital, Dundee

Carel L.G. Roux
Clinical Senior Lecturer/Honorary Consultant, Reader in Metabolic Medicine, Hammersmith Hospital, Imperial College London, London

Greg Sadler MD FRCS
Consultant Surgeon, John Radcliffe Hospital, Oxford

Sami M. Shimi MD BSc FRCS
Consultant upper GI Surgeon and Senior lecturer, Department of Surgery, Ninewells Hospital and Medical School, University of Dundee

Aileen Smith
Consultant Hepatologist, South West Liver Unit, Derriford Hospital, Plymouth

David Smith FRCS
Consultant Endocrine and General Surgeon, Ninewells Hospital and Medical School, Dundee

Elizabeth Starren MRCP
Academic Clinical Fellow in Respiratory Medicine Imperial College London

Martin Stotz MD FMH AFICM
Consultant, Intensive Care and Anaesthesia, Imperial College Healthcare NHS Trust, London, UK

Allan D. Struthers BSc MD FRCP FRCP(G) FRCP(E) DCH FESC
Head of Cardiovascular and Diabetes Medicine, University of Dundee Medical School, Ninewells Hospital, Dundee

Emile Tan MBBS MRCS
NIHR Clinical Lecturer in General Surgery, Department of Surgery and Cancer, Imperial College, London

Paris Tekkis BMedSci BMBS MD FRCS
Professor of Colorectal Surgery, Imperial College London and Academic Surgery Department, Chelsea and Westminster Hospital, London

Alistair M. Thompson BSc(Hons) MBChB MD FRCS(Ed)
Professor of Surgery and Molecular Oncology, Ninewells Hospital and Dundee Cancer Centre, University of Dundee, Dundee

Ian A. Zealley BSc MBChB FRCP FRCR
Consultant Radiologist, Ninewells Hospital, Dundee

PART I

General section

CHAPTER 1

Surgical biology and pathology

SIR ALFRED CUSCHIERI

Wound healing and repair

In earlier times the wound was assessed somewhat indirectly by studies of the breaking strength and collagen content. Today we look beyond these parameters to the behaviour of the individual cells in the wound since these are the prime movers in the healing process. Uneventful healing is dependent on their health and, in turn, this is determined by their microenvironment and a whole host of growth factors supplemented by growth-promoting and chemotactic substances produced primarily by platelets, lymphocyte cells and macrophages.

The elements of healing

Three distinct processes contribute to the process of wound healing.

Epithelialization

This is the process by which the surface covering of the wound is restored by a combination of cell migration and multiplication. The stimulus for epithelial repair is unknown. The loss of contact between cells undoubtedly plays a part. When an area is denuded of epithelium, the marginal cells divide and migrate across the bare area. The activity ceases when epithelial contact is re-established. There is also evidence that the loss of the epithelial cover is associated with a fall in the level of a local inhibitory hormone or chalone that is synthesized by the epithelial cells. Epithelialization proceeds most rapidly in a most highly oxygenated environment.

Contraction

This is the process by which the edges of an open wound gradually close together. It is a form of tissue migration that involves the entire thickness of the skin and subcutaneous tissues. It therefore proceeds most readily in areas where the skin is loose, such as the buttocks and the back of the neck. Contraction is the result of forces exerted by specialized fibroblasts in the wound. These cells have contractile elements in their cytoplasm known as myofibroblasts. The process of contraction is physiological and must be distinguished from the pathological process of scar contracture or cicatrization that causes distortion and limitation of movement.

Connective tissue formation

This is the process by which the main body of the wound is united. It plays a fundamental role in all but the most superficial injuries, and the strength of the wound following surgery is dependent on it. Connective tissue formation is the most important element of the three components, and many studies of wound healing are simply an examination of this component in isolation.

Types of healing

Although the elements of tissue repair are the same, open and closed wounds heal rather differently. When a surgical incision is closed with sutures or clips and heals without complication, it is said to heal by first intention. Union takes place by a combination of epithelialization and connective tissue formation. When an open wound is allowed to close naturally, union is accomplished by a combination of all three: wound contraction, connective tissue formation and epithelialization. This is known as healing by second intention or by granulation. These wounds heal from the bottom by abundant vascular connective tissue that has a granular appearance, hence the alternative name 'healing by granulation'.

Prior to the development of connective tissue, the closed wound is much more susceptible to infection than the open wound. For this reason, a heavily contaminated wound is often only partially closed. The deeper layers are secured but the subcutaneous tissues and the skin are left open. Once healing is established and granulation tissue has formed, the wound may be closed without fear of invasive infection developing. The technique is called delayed primary closure or secondary suture, and this type of healing is known as healing by third intention.

Phases of healing

In deeper wounds, the strength recovery is often an important attribute. Early studies demonstrated that this occurred in the phasic manner, typical of many biological processes. In the first few days, the wound has no recordable strength. Following this, strength increases rapidly. Finally, after a few weeks, the process slows down yet further increases in strength occur gradually. These features of strength recovery correlate well with observed changes in the wound, and healing is often considered as a three-phase event.

During the first few days, when the wound has no strength, little seems to be happening and this is called the lag phase. However, there is intense enzymic and leucocytic activity with breakdown and removal of devitalized tissue. A better term is the preparation phase, as the foundations for repair are being laid. During the next few weeks, the scene is dominated by the proliferation of cells and capillaries. Neutrophils and macrophages are prominent and fibroblasts lay down collagen in increasing amounts. In an open wound, this vascular fibrocellular tissue is recognized as granulation tissue. This is the phase of proliferation or fibroplasia. After a variable number of weeks, these wound activities slow down. Fibroblasts and capillaries are less evident and strength increases progressively. The third phase is one of maturation or differentiation that lasts for several months.

The organ of repair

During the phase of proliferation the new tissue in the wound can be thought of as a repair organ (Figure 1.1). Its delicate stroma and extensive capillary network provide both physical and nutritional support for its component cells. The function of this organ is connective tissue formation. The fibroblast is the key cell synthesizing collagen and intercellular ground substance. Its activities, together with those of the support macrophages, are dependent on a readily available supply of oxygen. As a result, cells and capillaries develop together as a unit and grow until the wound is filled. This new vascular connective tissue is most obvious in open wounds but it is also present between the edges of a healing, closed, incised wound. As time passes, the fibroblasts and the capillaries become much less prominent and a mature fibrous scar remains.

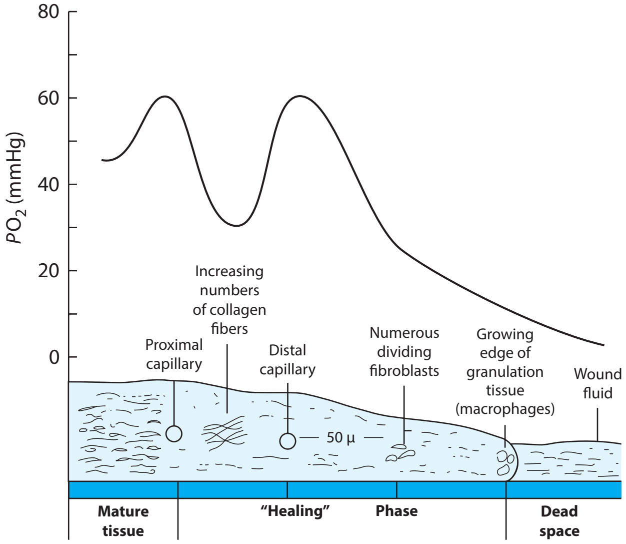

80
60
40
20
0
PO_2 (mmHg)
Increasing numbers of collagen fibers
Proximal capillary
Distal capillary
Numerous dividing fibroblasts
Growing edge of granulation tissue (macrophages)
Wound fluid
50 µ
Mature tissue
"Healing"
Phase
Dead space

Figure 1.1 Profile of a healing wound. It is a delicate system of cells and capillaries. New tissue grows from the 'vital' edge towards the central dead space. Direct measurements of PO_2 in this granulation tissue show a steady fall from the normal mature tissue level of around 45 mmHg to anoxic levels in the centre of the wound. Macrophages have a lower oxygen requirement than fibroblasts and are found at the free edge of the growing tissue. (After Silver, 1980.)

Wound strength

The recovery of strength has obvious clinical significance and has proved to be one of the most useful indications of the progress of repair. Early studies of wound-breaking strength showed that the apparently well-healed wound was still remarkably weak. Skin and scar tissue are complex viscoelastic materials that cannot be fully characterized unless tensile strength and extension (stretch) are recorded simultaneously. Plotted together, they describe a curve (load–extension) that reflects the ability of the scar to resist rupture (energy absorption, Figure 1.2). The findings are remarkably uniform, demonstrating only a 50–70% recovery of strength by the end of 6 months. It appears that total recovery is rarely achieved. Such low recordings need not cause alarm, for the absolute values are more than enough to meet the stresses imposed by everyday life.

Wound histology

Light microscopy shows a characteristic sequence of events (Figure 1.3). As time elapses after wounding, specific cell populations appear on the scene. Neutrophils predominate in the first day and monocytes peak about 24 hours later. By 5–6 days, fibroblasts are found in large numbers, and their presence is synchronous with the establishment of a microcirculation. Collagen is readily identified in increasing amounts after the fourth day.

A characteristic sequence of enzyme changes is also seen (Figure 1.4). When identified histochemically, these can be used to calculate the age of the wound in hours. Collagen is responsible for most of the strength of the wound, and scar weakness is associated with physical changes in the collagen. When the wound is examined by polarized light, normal collagen stands out as a clearly birefringent material. However, the wound scar does not exhibit this property during the first 6 months of healing. This lack of birefringence indicates a failure of organization at the molecular and small fibril level. Physical factors, e.g. fibre shape and weave, are important in determining the mechanical properties of skin and scar. These are best displayed by scanning electron microscopy. In unwounded skin, the collagen fibrils lie in well-organized bundles (Figure 1.5). In sutured wounds, the collagen fibrils lie in a relatively haphazard manner (Figure 1.6). As time passes, the collagen fibrils in the wound coalesce to form large irregular masses (Figure 1.7), remodelling being minimal. As yet, there is no evidence that the normal architectural network is ever restored.

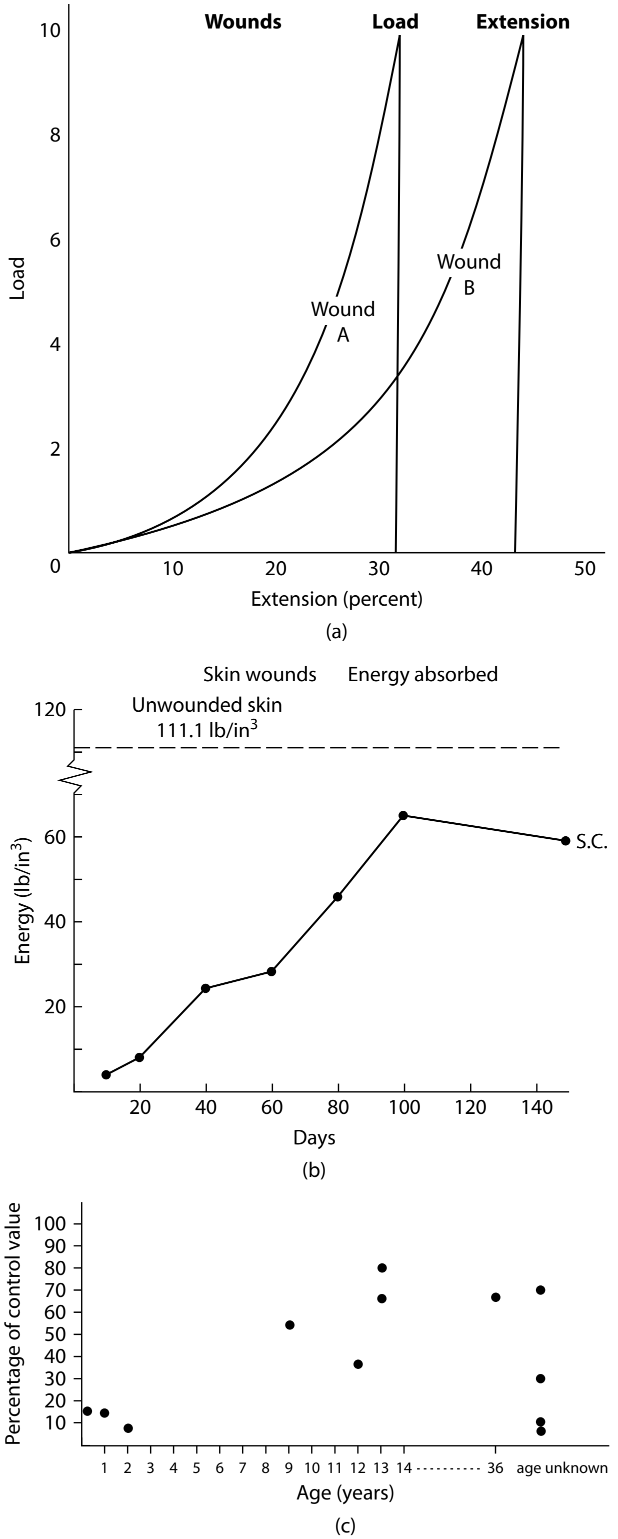


Figure 1.2 (a) Load–extension curves for wounds which break under the same load but differ in the degree of extension. Wound A is less pliable than wound B and is therefore easily ruptured. The ability to resist rupture (energy absorption) is measured by the area under the curve. (b) The ability of a wound to resist rupture expressed as its energy absorption. There is only a 50% recovery by 150 days. (From Forrester *et al.*, *J Surg Res* 1969;**9**:207–212, with permission.) (c) The tensile strength of human skin wounds expressed as a percentage of intact skin. In the first 2 year period, skin wounds are less than 20% of the control value, and even at 13 years there is still a marked weakness. (From Douglas *et al.*, *Br J Surg* 1969;**56**:219–222, with permission.)

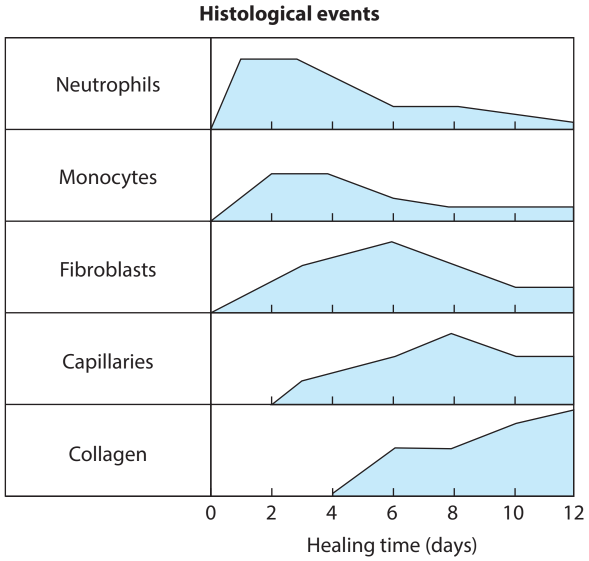


Figure 1.3 Characteristic sequence of events in the first few days of wound healing. Neutrophils and monocytes appear first. Collagen appears following the development of a functioning fibroblast–capillary system.

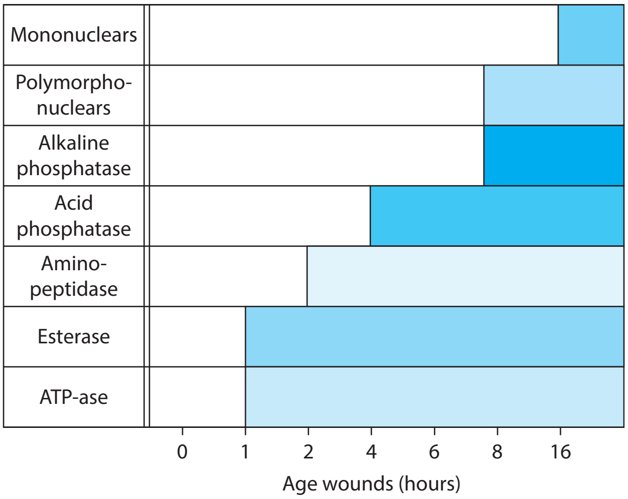


Figure 1.4 Schematic diagram showing the histochemical estimation of age of antemortem skin wounds. (From Rawkallio, *J Forensic Sci* 1972;**1**:3–16, with permission.)

Wound biochemistry

Of all the soft-tissue constituents of the body, only collagen has sufficient strength of its own to be responsible for the observed mechanical properties of unwounded tissue and firmly healed scar. The total amount of collagen rises rapidly in a healing wound and normal levels are usually attained within a few weeks. However, strength continues to increase long after the collagen content has returned to normal. Clearly, the quality of the collagen in the wound alters as time goes by.

Studies using radioactive tracer techniques have clarified the situation (Figure 1.8). The amount of collagen in the wound stabilizes after a few weeks but the rate of collagen synthesis and lysis remains high for considerably longer. This balance of synthesis and lysis may explain several healing defects. In keloid

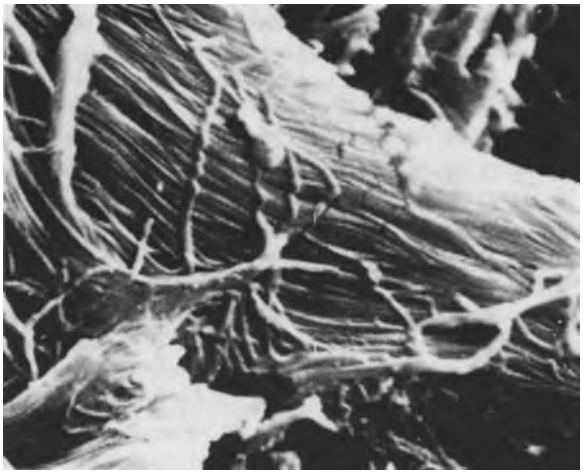


Figure 1.5 Scanning electron micrograph of part of a normal collagen fibre showing that it is made up of bundles of cross-banded fibrils (×9000). (From Forrester *et al.*, *Nature* 1969;**221**:373–374, with permission.)

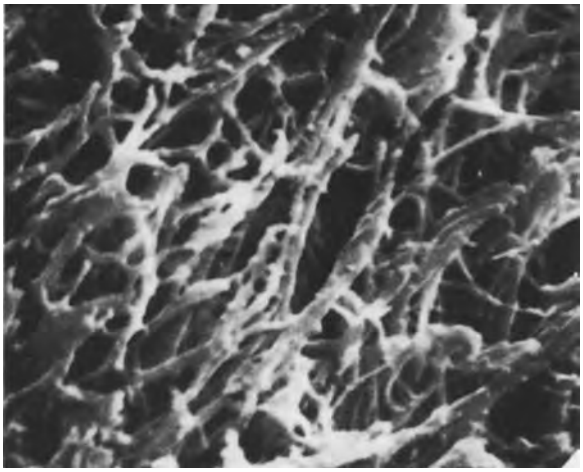


Figure 1.6 Scanning electron micrograph of a 10-day sutured wound showing the randomly orientated collagen fibrils. They show little tendency to aggregate. Cross-banding is not apparent (×9000). (From Forrester *et al.*, *Nature* 1969;**221**:373–374, with permission.)

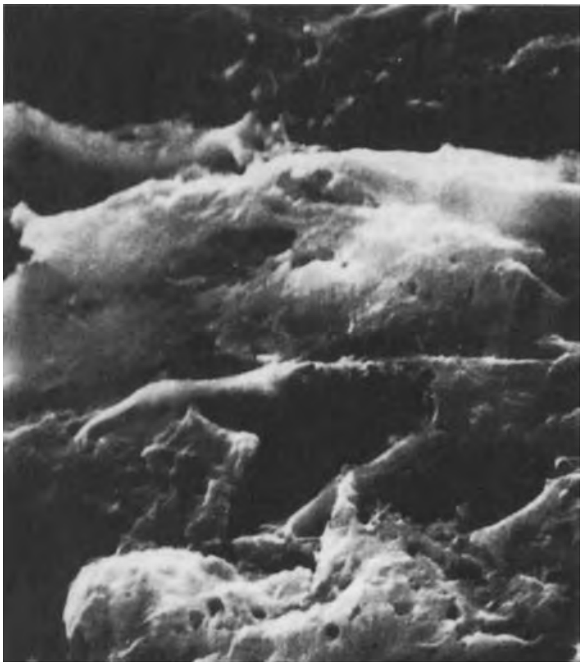


Figure 1.7 Scanning electron micrograph of a representative portion of a 100 day wound. The collagen fibrils have aggregated to form large collagen masses but normal fibre architecture has not been restored (×3150). (From Forrester *et al.*, *Nature* 1969;**221**:373–374, with permission.)

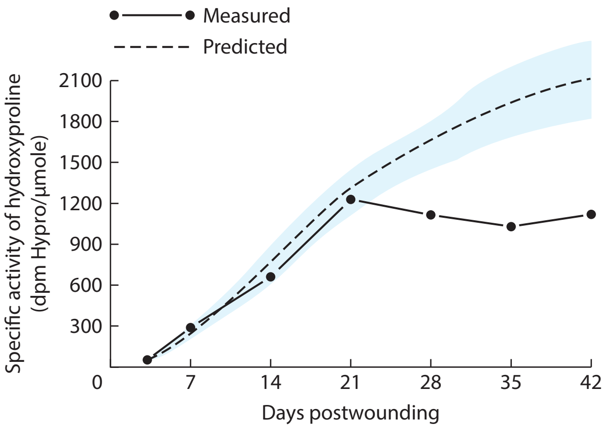


Figure 1.8 Comparison of scar collagen accumulation predicted from its rate of synthesis with that actually measured. Total collagen does not increase after 3 weeks, even though it continues to be synthesized and deposited at a rapid rate. Collagen is now being removed as quickly as it is formed (collagenolysis). The difference between the curves represents scar collagen turnover. (From Madden and Peacock, *Ann Surg* 1971;**174**:511–520, with permission.)

and hypertrophic scars, overproduction of collagen seems to be due to a relatively low rate of lysis. In scurvy, it is the synthesis of collagen that fails, and the wound weakens under the continued lytic process.

Collagen forms 30% of the total protein content of most animals. The collagen molecule is a rigid rod 300 nm long and 1.5 nm wide. Each molecule is composed of three polypeptide chains bound in a left-handed helix. The molecule itself is twisted the opposite way into a right-handed superhelix. The polypeptide chains of collagen are themselves remarkable. Over half of the molecule is composed of the three amino acids: glycine, proline and hydroxyproline. Both the carboxyl and amino ends of the molecule are non-helical and the entire structure is held together by hydrogen bonds. These bonds are relatively weak and the bulk of the strength of mature collagen is attributed to strong intermolecular and intramolecular covalent bonds. The individual amino acids are assembled in the endoplasmic reticulum of the fibroblast (Figure 1.9), beginning at the amino terminal and proceeding towards the carboxyl end.

A unique feature in collagen synthesis is that neither hydroxyproline nor hydroxylysine is incorporated directly into the collagen molecule. Instead, a proline-rich collagen precursor (protocollagen) is formed. Hydroxylation then proceeds under the influence of protocollagen hydroxylase. Requirements of the enzyme are oxygen, α-ketoglutarate, ferrous iron and ascorbic acid. Each of these may interfere with collagen metabolism but the only one of practical importance is ascorbic acid,

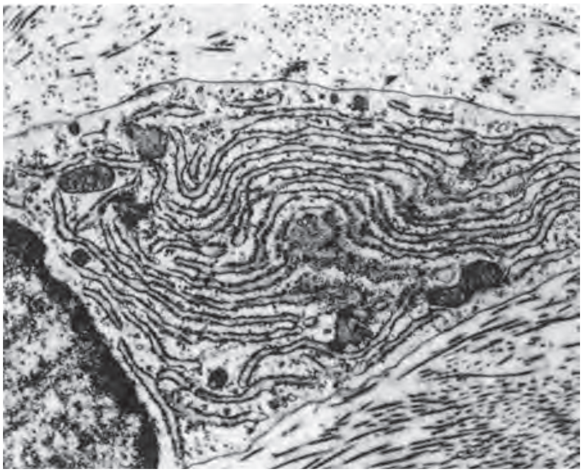


Figure 1.9 Electron micrograph of a normal fibroblast. Note the characteristic well-developed endoplasmic reticulum. The lining ribosomes, which are responsible for its rough appearance, are the active site of collagen synthesis. New collagen fibrils are rapidly excreted and are seen here surrounding the cell (×9000). (Courtesy of Professor Russell Ross, Seattle.)

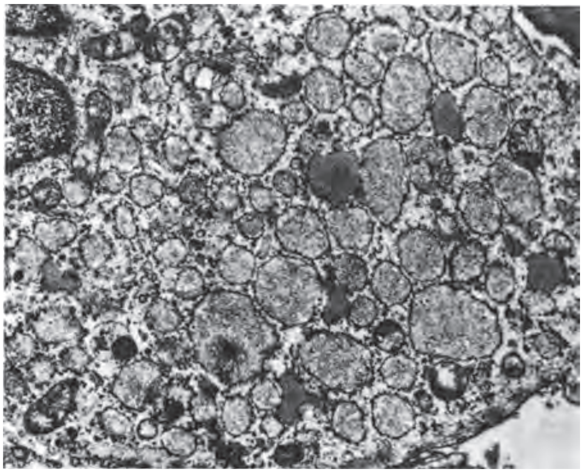


Figure 1.10 Electron micrograph of part of a scorbutic fibroblast. Note the typical distended endoplasmic reticulum. There is no sign of collagen but it will appear within 24 hours of providing ascorbic acid (×9000). (Courtesy of Professor Russell Ross, Seattle.)

deficiency of which delays collagen synthesis. The incompletely synthesized collagen cannot be excreted from the fibroblasts and distends their endoplasmic reticulum in a characteristic way (Figure 1.10). Several different genes direct synthesis and 13 distinct types of collagen have been identified in vertebrate tissues. Type I characterizes mature bone and skin. Type II is found in hyaline cartilage, and type III in cardiovascular structures, infant skin and the granulation tissue of healing skin wounds. Types IV and V are associated with basement membranes, and the others with a variety of tissues including cartilage and certain tumours. The significance of these different forms of collagen is not yet clear, but their existence does help to explain why cutaneous scar tissue behaves in a different way from the surrounding dermis. Scar collagen contains types I and III and differs in its degree of hydroxylation of lysine and glycosylation of hydroxylysine. The cross-linking pattern is also characteristic.

All connective tissues contain varying amounts of ground substance. This amorphous matrix between the cells and fibres contains protein–glycosaminoglycan complexes called proteoglycans. The fibroblast synthesizes collagen, glycosaminoglycans and fibronectin. The last component of the matrix is a large glycoprotein with important influences on both intercellular adhesion and cell-to-matrix adhesion. The functions of the proteoglycans and fibronectin are incompletely understood, but they appear to play a part in the organization and precipitation of collagen fibres. The ground substance has, in addition, important effects on the mechanical properties of mature tissue. The wound is a fibre–gel–fluid system and the mechanical properties of this complex material differ significantly from those of fibrous tissue alone.

Factors affecting healing

A number of factors influence the rate of wound healing. However, some have unpredictable effects. Thus, jaundice and uraemia adversely affect wound healing in animals but not in humans. The important factors influencing healing in clinical practice are age nutrition, vascular supply, sepsis, oxygen and wound dressing.

Age

Healing proceeds more rapidly in the young provided they are well nourished. The increased vigour of repair may explain why hypertrophic scars and keloids are more common in early life.

Nutrition

To a certain extent, wounds do not heal well in the debilitated and malnourished. However, several studies have documented the biological priority of healing wounds. Thus, patients have to be very severely protein depleted before healing is impaired. Ascorbic acid is required for the synthesis and maintenance of collagen. Following injury, body stores are rapidly depleted and a scorbutic state may be induced. When this happens, collagen synthesis is impaired with delayed healing. In older wounds in which collagen turnover is still active, scars have been known to reopen. Healing is also delayed by zinc deficiency.

Vascularity

Wounds heal well in areas such as the face where the blood supply is good, and vice versa. The most striking examples are found in ischaemic vascular disease of the lower limb. Both wound healing and the overall metabolic response to trauma are optimal when the environmental temperature is raised to 30°C. The combination of increased blood flow and warmth following sympathectomy has been shown to improve healing in patients with peripheral vascular disease.

A minimal inflammatory stimulus is required for healing to progress normally. If anti-inflammatory drugs, e.g. cortisone, are administered in the first few days after wounding, healing is likely to be delayed. Once healing is established, cortisone does not appear to interfere with the healing process. In practice, wounds do heal in patients receiving long-term steroid therapy, but the process is slow and more susceptible to complications.

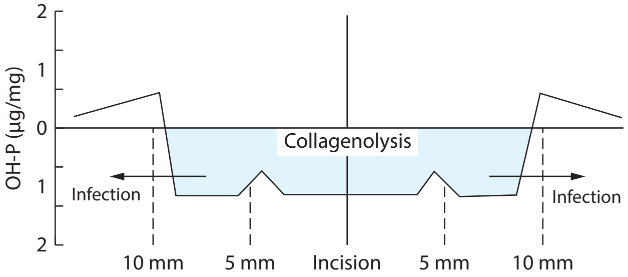


Figure 1.11 The chemically active zone of an incised wound extends for at least 5 mm on either side of the wound. Collagen lysis is prominent in the first week and is more marked when infection is present. The zero line is the concentration in normal abdominal wall. (After Adamsons *et al.*, *Surg Gynecol Obstet* 1966;**123:**515–521.)

Sepsis

Local infection is perhaps the most important cause of delayed wound healing and dehiscence. Collagen synthesis is depressed and collagenolysis increased, thereby enhancing the softening of the wound edges (Figure 1.11). This adversely affects the strength of the repair and makes cutting out of the sutures more likely.

Oxygen

Oxygen is crucial for wound healing. Direct examination of the granulation tissue growing into a wound chamber shows a number of regular features. The PO_2 falls steadily from the normal mature tissue level of around 45 mmHg (6.0 kPa) to levels close to zero in the centre of the wound. Fibroblastic activity is maximal up to 50–80 µm away from the nearest normally perfused capillary. At this point, PO_2 levels between 10 and 20 mmHg (1.3–2.6 kPa) are regularly recorded. Macrophages have lower oxygen requirements than fibroblasts and are found at the free edge of the growing granulation tissue. Even in these areas of very low oxygen tension, they are still able to ingest bacteria, but there is uncertainty about their ability to kill ingested organisms. Increased oxygen uptake is invariably associated with bactericidal activity since the process is mediated by the peroxidase system.

The delivery of nutrient oxygen to the wound is impaired by a number of local factors, such as tissue trauma and tight suturing techniques. More serious problems arise when wound capillary perfusion is impaired by systemic disorders. By far the most serious of these is the capillary shutdown associated with hypovolaemia. The wound, together with the splanchnic and cutaneous circulation, is the first to shut down in an attempt to maintain circulation to the vital organs. Similar effects are observed in states of increased blood viscosity and cardiopulmonary decompensation. Finally, there is evidence that increasing the oxygen supply to a wound induces greater collagen production (Figure 1.12), although it does not appear to affect the overall rate of healing.

Wound dressings

The undisturbed wound heals best, and dressings may therefore impair healing, especially of open wounds, by damaging the delicate new cells and capillaries on the wound surface. These wounds have to be packed with non-adherent materials that permit surface oxygenation and do not encourage bacterial overgrowth.

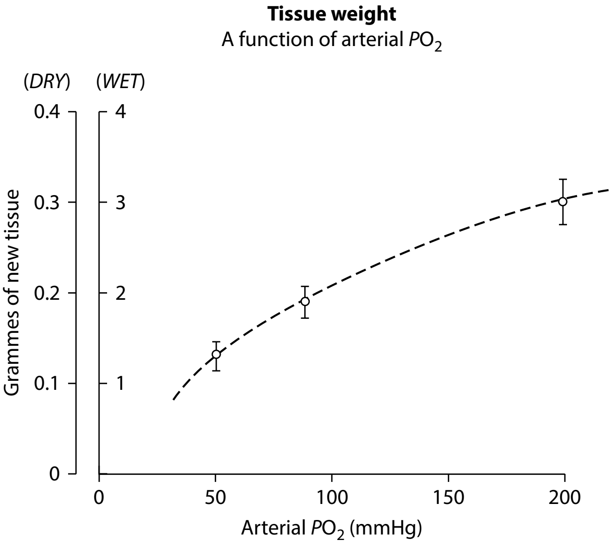


Figure 1.12 The amount of new tissue formed in a wound is considerably greater when arterial PO_2 is increased by changing the ambient oxygen from 14–20% to 45% for 25 days. (From Hunt, *Trauma* 1970;**10:**1001–1009, with permission.)

Wound failure

Although healing is a unified response to injury, failures tend to present in three quite distinct ways. Acute failures are wound infection and dehiscence. Chronic failure is the condition of pathological fibrosis resulting from the overproduction of scar tissue (hypertrophic scar, keloid).

Acute failures

Wound infection

This is the most common and troublesome disorder of wound healing. A primarily closed wound has no resistance at all to bacteria contaminating the surface during the first 6 hours (Figure 1.13). After this time, it becomes increasingly difficult to infect the wound, until at 5 days it is as resistant as the surrounding skin. Thus, an occlusive dressing is advisable only during the first few days unless there is an obvious nearby source of contamination, e.g. colostomy. *The main source of wound infection is endogenous from the patient's own bacteria at the time of surgery*.

The factors predisposing to wound infection include:

- local trauma from excessive retraction, extensive electrocoagulation, defective haemostasis
- the presence of foreign material; a single piece of sterile silk suture material doubles the chance of a contaminated wound becoming infected (Figure 1.14)
- diminished perfusion.

Wound dehiscence

This is always a catastrophe. At the least, the patient requires a second operation and hospital stay is prolonged. In specific

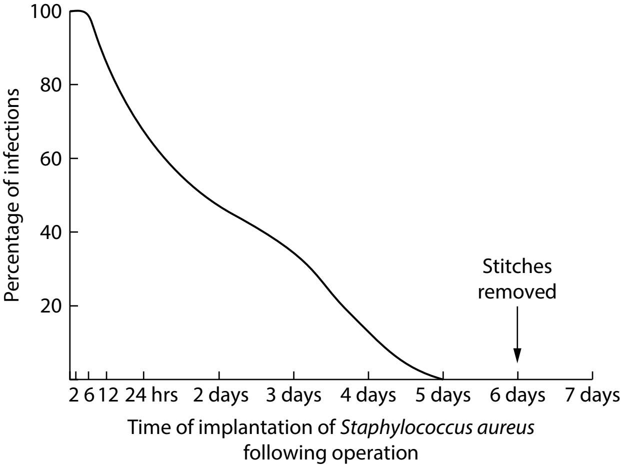


Figure 1.13 Vulnerability of a healing incised wound to surface contamination with micro-organisms. During the first 6 hours it has no resistance. Thereafter, it becomes increasingly resistant to invasion, and by 5 days it is as resistant as normal skin. (After DuMortier, *Surg Gynecol Obstet* 1933;**56:**762–766.)

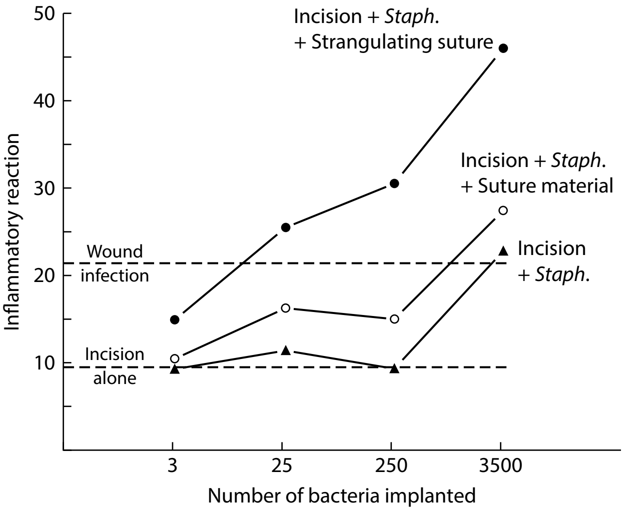


Figure 1.14 The presence of extraneous material in a wound enhances the likelihood of infection developing. The presence of one tied silk suture doubles the chance of a contaminated wound becoming infected. (After Howe, *Surg Gynecol Obstet* 1966;**123:**507–514.)

situations, particularly in cardiac and vascular surgery, it may prove fatal. In a high proportion of cases, wound infection precedes and determines the result. Wound dehiscence after laparotomy is a serious complication, occurs in 0.25–3.0% of laparotomies and has a reported mortality of 20%. Risk factors include anaemia [haemoglobin (Hb) <10 g], hypoalbuminaemia, malnutrition, malignancy, jaundice (debatable), obesity [body mass index (BMI) >35], sepsis, diabetes, male sex and elderly patients (>70 years). There is also an increased incidence after specific operations, e.g. emergency laparotomy and total colectomy. Although preoperative chemotherapy and radiotherapy are reported as risk factors, the evidence for this is by no means certain. Wound dehiscence is commoner with midline vertical wounds and when wounds are closed with interrupted sutures as opposed to continuous closure. The highest incidence is encountered in elderly patients with chronic obstructive airways disease, and in some reported series these patients account for 70%.

Ordinary, the condition becomes clinically manifest on days 7–9, although occasionaly wound dehiscence can occur after discharge from hospital. It is usually preceded by seepage of blood-stained serum from the wound for some hours before the actual dehiscence; otherwise, the diagnosis is obvious, with loops of intestine appearing in the depths of the wound or even prolapsing when the patient coughs or strains. Patients often develop some degree of hypotension.

The immediate management is by a sterile dressing and bandaging, therapeutic nasogastric decompression, intravenous fluid therapy and intravenous broad-spectrum antibiotics followed by immediate surgery unless the patient requires resuscitation. During operative closure, some advocate the use of deep tension sutures, although the value of these remains controversial.

Wound fibrosis

Abnormally contracting tissue is often a late consequence of injury or inflammatory disease and features a whole range of chronic fibrotic processes, from simple adhesions in peritoneum and tendon sheaths to interstitial fibrosis. Other troublesome examples are benign oesophageal stricture, mitral stenosis and hepatic cirrhosis. In primary closed wounds pathological fibrosis results in either a hypertrophied scar or keloid. Wound scarring is also increased in wounds that become infected before they heal. Attempts have been made to control this pathological fibrosis by specific antifibrotic treatment of the scar (β-aminopropio-nitrile, penicillamine, steroids and zinc) and non-specific methods aimed at diminishing the inflammatory process that precedes fibroblast activation.

The present methods of managing hypertrophic scars and keloids are imprecise and rely on compression, surgical excision and judicious use of corticosteroids (triamcinolone).

Blood and haemostasis

Although the cardinal function of blood and an effective circulation is the provision of adequate tissue perfusion for the supply of oxygen and for export of waste products, the cellular component and the plasma fraction of blood subserve several other functions, some of which involve the participation of the monocyte–macrophage system.

Blood groups and histocompatibility antigens

Blood groups

The most important is the ABO system. The surface antigens A and B are coupled to ceramide, a lipid component of the cell membrane of the erythrocyte. Both are derived from substance H as a result of the activity of two specific transferases coded by the respective allelic genes A and B on chromosome 9. Blood group O individuals do not possess either of these transferase enzymes and thus carry substance H unchanged. Antibodies to antigen A or B are present in the plasma when the respective antigen is

Table 1.1 ABO blood groups

Phenotype (group)	Erythrocyte genotype	Antigens	Antibodies	Incidence (%)
O	OO	H	Anti-A, anti-B	47
A	AA/AO	A	Anti-B	42
B	BB/BO	B	Anti-A	9
AB	AB	AB	None	3

absent from the red cells. Thus, group A individuals have anti-B; group B individuals have anti-A; group O individuals have both anti-A and anti-B; and group AB individuals do not have any. Although group O individuals are often referred to as 'universal donors', they should always be screened for anti-A and anti-B before their blood is administered to group A and B patients. These natural antibodies, which belong to the immunoglobulin (Ig) M class, are haemolysins and are thought to arise as a result of immunization by gut bacterial antigens closely allied in composition to the A and B antigens. The characteristics of the ABO blood group system are outlined in Table 1.1.

Of the other blood group systems (rhesus, Kelly, Duffy and Kidd), the most clinically relevant is the rhesus (Rh) system. This is inherited by a single complex gene on chromosome 1, which gives various combinations of C or c, D or d, and E or e. An Rh-negative mother (genotype = dd) will be immunized by her fetus if this is Rh positive (genotype = DD), usually as a result of placental bleeding during delivery of the firstborn, with the production of anti-D haemolysins. During subsequent pregnancies, these IgG anti-D antibodies cross the placenta and cause haemolytic disease of the newborn if the fetus is Rh positive. This disease has been virtually eliminated as a result of treatment of the mother with preformed anti-D IgG at the time of birth of her first baby to haemolyse any Rh-positive cells from the baby that reach the maternal circulation.

Leucocyte and platelet antigens

Antigens on the surface of leucocytes and platelets may be cell specific or present on other cells of the body. The latter are known as shared antigens and the most important group in this category is the human leucocyte antigen (HLA) system. It is also referred to as the major histocompatibility complex (MHC) and provides the mechanism by which the immune system recognizes self. These antigens belong to two classes (I and II) and are of great importance in organ transplantation.

Erythropoiesis and function of the erythrocyte

Erythropoiesis

Normal physiological function is dependent on the maintenance of a constant mass of circulating red blood cells (approximately 309×10^9 dL), the primary function of which is the supply of oxygen and removal of CO_2 essential for cellular respiration, continued viability and metabolic activity.

The circulating red cell mass is controlled by balanced replacement of erythrocyte loss due to senescence (normal life span of 120 days) with production (erythropoiesis) in the bone marrow and liver (fetus). Effete erythrocytes are removed by the monocyte–macrophage system, predominantly in the spleen. Marrow hyperplasia and increased activity are encountered in all conditions associated with abnormal red cell destruction, i.e. haemolytic anaemias.

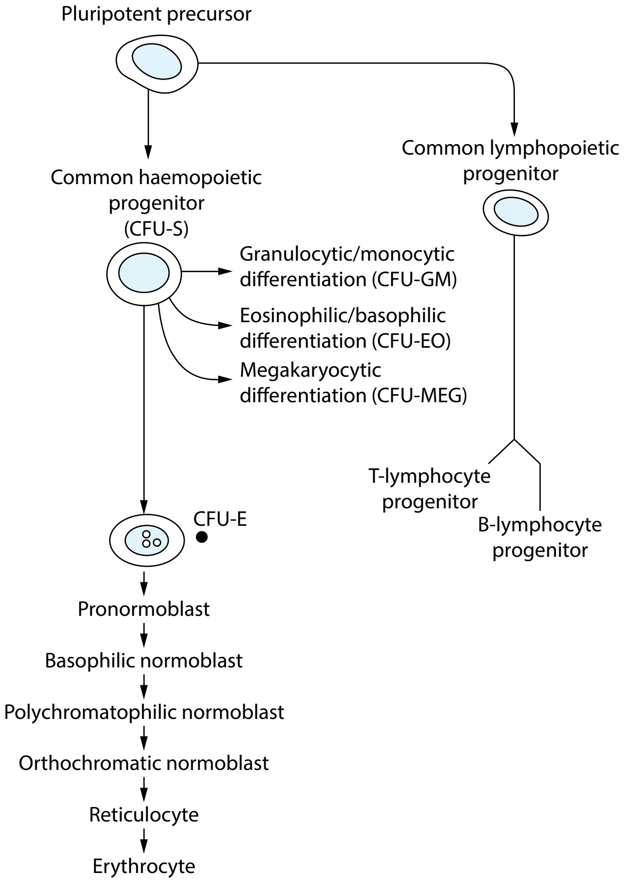

Pluripotent precursor
Common haemopoietic progenitor (CFU-S)
Common lymphopoietic progenitor
Granulocytic/monocytic differentiation (CFU-GM)
Eosinophilic/basophilic differentiation (CFU-EO)
Megakaryocytic differentiation (CFU-MEG)
T-lymphocyte progenitor
B-lymphocyte progenitor
CFU-E
Pronormoblast
Basophilic normoblast
Polychromatophilic normoblast
Orthochromatic normoblast
Reticulocyte
Erythrocyte

Figure 1.15 Schematic representation of normal erythropoiesis.

Erythrocytes originate from a pluripotential cell via the common haemopoietic progenitor [spleen colony-forming unit (CFU-S)]. Under the influence of erythropoietin, some of these CFU-S differentiate into several erythroid colony-forming units (CFU-E), which then differentiate further and mature into red blood corpuscles (Figure 1.15). Erythropoiesis is predominantly regulated by the oxygen tension in the renal blood, which regulates the release of proerythropoietin that is then activated to erythropoietin in the plasma. Low renal oxygen tension leads to increased release of erythropoietic factor by the kidneys and thus enhance erythropoiesis. Recombinant erythropoietin is now available and is used predominantly in the treatment of anaemia associated with chronic renal disease. In addition to erythropoietin, several other growth factors, hormones (growth, thyroid, insulin, androgens), vitamins (B_6, B_{12}), folate, iron and inhibitors (T-lymphocyte factors) are involved in this multistage process. The earliest microscopically recognizable erythrocyte precursor in the bone marrow is the pronormoblast (rubriblast, proerythroblast).

Red cell function

The major fuel supply to the erythrocyte is glucose that is metabolized largely through anaerobic glycolysis (Embden–Meyerhof pathway) and, to a much lesser extent, via the pentose phosphate pathway. Anaerobic glycolysis yields ATP (to meet cellular energy requirements), NADH and 2,3-diphosphoglycerate (2,3-DPG), which influences the affinity of Hb for oxygen. The pentose phosphate pathway produces NADPH. The NADH/NADPH system is necessary to protect against oxidative damage to the Hb.

The main function of the erythrocyte is the transport of oxygen and CO_2. Hb is essential for the carriage of oxygen, as this is only slightly soluble in water. For this reason, at normal atmospheric pressure, only a very small amount of oxygen (approximately 3.0 mL/L) is in solution in the plasma. Thus, the bulk of the oxygen requirement (250 mL of O_2/min at rest) is carried by the erythrocytes. The haemoglobin molecule consists of two α- and two β-chains, with each chain being folded over a haem group. This configuration enables the binding of four molecules of oxygen (one per chain). Thus, oxyhaemoglobin is represented as $Hb(O_2)_4$. At normal atmospheric pressure, the partial pressure of inspired air in the alveoli normally exceeds 100 mmHg (13.3 kPa), whereas the partial pressure in the pulmonary arterial blood reaching the alveolar capillary network is much lower (40 mmHg; 5.3 kPa). This pressure gradient permits rapid diffusion across the alveolocapillary membrane to the plasma for uptake by the Hb in the erythrocytes. This results in 95% saturation of the Hb with oxygen. The PO_2 in the blood (and thus the amount of oxygen in solution in the plasma) can be increased by oxygen therapy and especially by hyperbaric therapy without affecting the amount of oxygen carried by Hb (because this is fully saturated). Hence, breathing pure oxygen instead of air (which contains only 21% oxygen) can raise the arterial PO_2 to 600 mmHg (80 kPa), increasing the amount of oxygen in solution accordingly. Much higher partial pressures of oxygen in the arterial blood are obtained by hyperbaric therapy.

At the blood–tissue interface, oxyhaemoglobin releases oxygen (dissociates) as a result of complex changes in its chemical and structural configuration. Primarily, the extent of dissociation (and conversely the percentage saturation) of Hb is dependent on the PO_2 in the interstitial fluid. This is outlined by the sigmoid oxyhaemoglobin dissociation curve (Figure 1.16). This dissociation curve can be shifted to the left (Figure 1.16), reflecting an increased affinity of Hb for oxygen. This is deleterious as oxygen is less readily released to the tissues. A shift to the left can be brought about by:

- low PCO_2 (hyperventilation)
- high pH (alkalosis)
- hypothermia
- reduced content of 2,3-DPG (stored blood).

By contrast, a shift to the right implies decreased affinity of Hb for oxygen and thus increased delivery of oxygen to the tissues. It is encountered as a physiological compensatory mechanism in pyrexia (from any cause) acidosis and hypercapnia. The relation between shifts in the oxyhaemoglobin and pH changes is known as the Bohr effect.

Simultaneously with the release of oxygen to the tissues, CO_2 diffuses into the erythrocytes. Thus, the essential ongoing first reaction at the blood–tissue interface is:

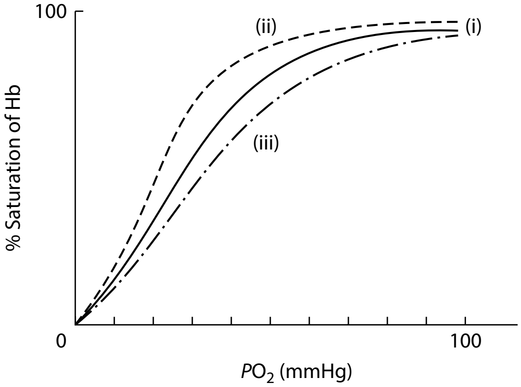


Figure 1.16 Oxyhaemoglobin dissociation curves: (i) normal curve; (ii) dissociation curve shifted to the left reflecting an increased affinity of Hb for oxygen, which is deleterious as oxygen is less readily available to the tissues; (iii) a shift to the right implies a decreased affinity of Hb for oxygen and thus increased delivery of oxygen to the tissues. It is encountered as a physiological compensatory mechanism in pyrexia, acidosis and hypercapnia. The relation between shifts of the oxyhaemoglobin and pH changes is known as the Bohr effect.

$$Hb(O_2)_4^- + H_2O + CO_2 \leftrightarrow Hb^- + H_2CO_3 + 4O_2$$

This reaction is catalysed by the enzyme carbonic anhydrase. Within the physiological pH range, most of the carbonic acid (H_2CO_3) dissociates:

$$H_2CO_3 \leftrightarrow H^+ + HCO_3^-$$

The reduced Hb^- takes up the liberated H^+ to form HHb (isohydric exchange) and the HCO_3^- diffuses out of the cell in exchange for Cl^- (chloride shift). By this process, the pH of venous blood is not significantly altered despite carriage of CO_2. These reactions are reversed at the alveolar capillary membrane in the lungs with the release and escape of CO_2 into the alveolar air and oxygenation of the reduced haemoglobin.

Abnormal haemoglobins

Abnormal affinity for oxygen, increased red cell aggregation or haemolytic episodes can be caused by a variety of disorders of haemoglobin (congenital and acquired). These include hereditary genetic coding disorders characterized by either qualitative (haemoglobinopathies, e.g. HbSS in sickle cell disease) or quantitative [e.g. reduction or absence (thalassaemia) of the globin polypeptide chains] changes. In addition, there are abnormal non-functioning derivatives of haemoglobin. The latter are often acquired and include methaemoglobin (drug induced, e.g. sulphonamides, severe acute pancreatitis), sulphhaemoglobin (drug induced) and carboxyhaemoglobin (CO poisoning, smoking).

Monocyte–macrophage system

Also known as the reticuloendothelial system, this consists of the circulating monocytes and fixed macrophages in lymph nodes, spleen, lungs and the liver, which is the largest component. It is now generally believed that the fixed tissue macrophages are derived from circulating monocytes. The main functions of the monocyte–macrophage system are as follows:

- Phagocytosis and degradation by lysosomal enzymes of particulate matter and bacteria; in this respect, the Küpffer cells and the macrophages in the lung filter bacteria from the bloodstream.
- Removal of senescent red blood cells, especially in the spleen and, to a lesser extent, the liver.
- Antigen presentation: in lymphoid tissue macrophages ingest and break down large antigens into soluble fragments, which are then presented on the surface to β-cells and MHC class II-bearing T-cells. Specialized antigen-presenting macrophages in the lymphoid follicles have dendritic processes that capture immune complexes, retain the antigen from these complexes for long periods and present it to numerous B-lymphocytes, including memory cells. Antigen presentation to T-cells in the paracortical areas of lymph nodes is performed by a different type of macrophage, possibly derived from the dendritic Langerhans cells of the skin
- Storage of iron involving intracellular binding to apoferritin.
- Synthesis of coagulant factors (e.g. fibrinogen), proteinases (e.g. α_2-macroglobulin), cytokines [interleukin (IL) 1], lymphokines (e.g. as colony-stimulating factors that regulate proliferation and differentiation of granulocytes, macrophage precursors and erythropoiesis), erythropoietin, etc.

The monocyte–macrophage system is implicated in various hereditary storage disorders resulting from specific enzyme deficiencies, e.g. various mucopolysaccharidoses and several lipidoses, the most common of which is Gaucher disease.

Function of plasma proteins

The plasma proteins are largely responsible for the oncotic pressure that is crucial to the exchange of fluid across the capillary membrane. In addition, they have other important functions. These include blood clotting, anticoagulation and inhibition of fibrinolysis, transport, buffering of acid metabolites, acute phase response and defence against microbial invasion through the complement–antibody response assisted by the immune cellular response.

Oncotic pressure

The plasma proteins, particularly the albumin fraction, are responsible for maintenance of the circulating blood volume. The effect is mediated by the oncotic pressure that they exert. This reverses the fluid flux from the interstitial space to the intravascular compartment at the venular end of the capillary where the effective oncotic pressure of the plasma proteins exceeds the capillary hydrostatic pressure. This delicately balanced mechanism is disturbed if the plasma protein level is decreased (hypoalbuminaemia) or the hydrostatic pressure in the capillary bed is elevated through congestion (e.g. right heart failure, circulatory overload, varicose veins) or the permeability of the capillary membrane is increased (sepsis, inflammation). All of these events result in expansion of the interstitial fluid compartment with oedema formation.

The normal bidirectional flow of fluid across the capillary membrane also supports the exchange of nutrients and waste products between the capillary blood and the interstitial fluid. The actual exchange is, however, independent of net flow and is governed by the movement of solutes along concentration gradients.

Anticoagulation

A substantial percentage of the globulin fraction of the plasma proteins (approximately 20%) is made up of agents that inhibit the activity of activated clotting factors and fibrinolytic enzymes (Box 1.1). Essentially, these are antiproteinases, although some have other specific physiological roles. As a group they limit and localize thrombosis, fibrinolysis and inflammatory reactions. Hereditary deficiency states may result in recurrent thrombosis, bleeding tendency (α_2-antiplasmin, plasminogen activator inhibitor 1), pulmonary emphysema (α_1-antitrypsin) or hereditary angioneurotic oedema (C1 inactivator).

Transport

The plasma proteins act as an efficient transport system for a variety of endogenous and exogenous substances: hormones, bilirubin, iron (transferrin), copper (caeruloplasmin), vitamins (retinol-binding protein), drugs, etc. In some instances, binding also serves to reduce the innate toxicity of material (e.g. binding of unconjugated bilirubin to albumin). In general, when a substance is bound to the plasma proteins it is inactive. Thus, release from the bound state is necessary for the physiological effect of the hormones on target organs.

Buffer action

Plasma proteins, like all other body proteins, have a high content of weakly acidic and basic groups and thereby contribute to the general buffering mechanism of blood. CO_2 generated by cellular respiration dissolves in the aqueous component of tissue fluid and blood to form carbonic acid (H_2CO_3), which readily dissociates at the normal blood pH range to hydrogen ions (H^+) and bicarbonate (HCO_3^-). Thus, carriage of CO_2 by the venous blood tends to lower pH by increasing the H^+ concentration. However, in the physiological healthy state, the pH of venous blood is only marginally lower than that of arterial blood. This is in part the result of the buffer base supplied by plasma phosphates and plasma proteins. However, oxyhaemoglobin–haemoglobin provides the main buffering of H^+ generated by CO_2 carriage through the process of isohydric exchange.

BOX 1.1 Natural plasma inhibitors

- Antithrombin III (heparin cofactor)
- Heparin cofactor II
- α_2-macroglobulin
- α_1-antitrypsin
- α_2-antiplasmin (α_2-plasmin inhibitor)
- C1 inactivator (C1 esterase inhibitor)
- Extrinsic pathway inhibitor
- Activated protein C inhibitor

Defence against microbial infection: complement system, immunoglobulins and acute phase reactants

Complement system

As with blood coagulation, the complement system consists of a series of enzyme proteins that act in a sequential manner to produce molecular fragments and complexes, which subserve important functions in the host's defence against microbes. Thus, coating of bacteria with fragment C3b facilitates their adherence to phagocytes that have cell-surface receptors for this component. Small peptide complement fragments (C3a and C5a) have intrinsic biological activity and either trigger the release of preformed mediators from mast cells and basophils (histamine, tryptase, chemotactic factors for polymorphs and eosinophils and platelet-activating factor) or activate phospholipase, which releases arachidonic acid from the phospholipids of the cell membrane of the mast cells. The arachidonic acid is then converted to leukotrienes (lipo-oxygenase pathway) or to prostaglandins and thromboxanes (cyclo-oxygenase route). Leukotrienes induce vasodilatation, bronchoconstriction and chemotaxis, whereas prostaglandins and thromboxanes cause platelet aggregation and vasodilatation. Collectively, these agents provide the final common motor pathway for the acute inflammatory response (vasodilatation, exudation of plasma proteins and accumulation of polymorphs). Finally, complement, by the formation of the membrane attack complexes (MACs), consisting of several complement components (C5b, C6–C9) on the surface of the invading organism, increases the permeability of its cell membrane. Thereafter, water and electrolytes enter the cell attracted by osmotic forces in such quantities as to induce cell disruption and lysis.

The complement system is activated by two mechanisms: the alternative and the classic pathways. Some micro-organisms can bind the C3 cleaving (convertase) enzyme directly to their surface polysaccharide molecules and release it from inhibition. The activated enzyme (C3bBb) then generates large amounts of C3b cleavage products from C3 that are deposited on the surface of the organisms. This direct activation occurring in the absence of antibody constitutes the alternative pathway of complement activation. By contrast, the classic pathway is initiated by binding of antibody to the surface antigens of the invading organism. This binds to and activates the C1q component of complement, setting off a multistep cascade reaction involving several complement components and leading to the formation of an activated enzyme (C4b2b) that has C3 convertase activity. Thus, the two pathways converge at the cleavage of the C3 component, which is pivotal to the complement system.

Antibodies

The body has a vast army of B-lymphocytes bearing different antibodies with specific recognition sites for foreign antigenic material. Thus, when an organism gains access, lymphocytes with the appropriate surface antibody 'dock' with the antigen and thereafter proliferate to form a clone of plasma cells programmed to produce large amounts of the required antibody needed to eliminate the invading organism. This process of proliferation of a specific line of B-lymphocytes is known as clonal selection and forms the basis of the acquired immune response. This encounter with a specific pathogen is permanently coded in some of these lymphocytes, which persist as memory cells. When compared with the primary response described above, these cells are capable of mounting a more rapid and vigorous antibody response (higher antibody titres) in the event of a second encounter with the same pathogen (secondary response).

Acute phase reactants

This is a collective term for a group of plasma proteins, the concentration of which rises dramatically in acute inflammation and trauma (Box 1.2). The group includes a wide range of proteins, which normally subserve specific physiological functions: ferritin, fibrinogen, C-reactive protein, caeruloplasmin, α_1-antitrypsin, α_2-macroglobulin, C9, haptoglobin, etc. The most clinically useful in the context of acute illness is C-reactive protein, which is synthesized and secreted by the liver. Endotoxin (during an infection) stimulates the release of IL-1 from the Küpffer cells, which leads to enhanced hepatic synthesis of the protein. In turn, C-reactive protein binds to some bacteria and thereby activates complement. Hence, it performs a useful opsonizing effect, i.e. encourages adherence to phagocytes. In surgical practice, circulating levels of C-reactive protein are often used as an index of severity of the illness, e.g. acute pancreatitis.

Haemostasis

In health, perfusion of tissues and organs is dependent on vascular patency, adequate perfusion pressure, normal blood rheology and interacting physiological mechanisms that prevent intravascular coagulation. At the same time, the body can mount a rapid co-ordinated response aimed at achieving efficient haemostasis in the event of vascular injury. Disorders of the mechanisms responsible for maintaining the fluid nature of blood result in the hypercoaguable state with a tendency to recurrent thrombosis. By contrast, defects in the haemostatic mechanism are manifested by bleeding, which may be spontaneous or occasioned by trivial injury.

Physiological control mechanisms against intravascular coagulation

These are outlined in Box 1.3. An adequate blood flow is crucial to the prevention of intravascular thrombosis as it dilutes and clears activated coagulants away from the site of the injury. In clinical practice, this is well exemplified by venous thrombosis, in which the combination of a hypercoaguable state and venous stasis is required to initiate this pathological event.

BOX 1.2 Disorders inducing a major acute phase response

- Bacterial infections
- Rheumatic disease: rheumatoid arthritis, seronegative spondarthritis
- Vasculitis
- Crohn's disease
- Trauma including burns and surgery
- Malignancy

BOX 1.3 Mechanisms responsible for preventing intravascular coagulation in the healthy state

- Blood flow
- Naturally occurring plasma inhibitors
- Heparin sulphate–antithrombin III mechanism
- Protein C–thrombomodulin–protein S mechanism
- Fibrinolysis
- Endothelial prostacyclin
- Hepatic clearance of activated factors

In addition to the plasma protein inhibitors described previously, two important mechanisms play a key role in the prevention of intravascular coagulation. These are the heparin sulphate–antithrombin III (AT-III) mechanism and the protein C pathway AT-III is a protease inhibitor synthesized and secreted by the liver. Native AT-III neutralizes the activity of thrombin and other clotting factors rather slowly but becomes a rapid and potent inhibitor in the presence of mucopolysaccharides such as heparin sulphate elaborated by the vascular endothelium. The physiological role of the heparin sulphate–AT-III mechanism is therefore to neutralize activated clotting factors on the vascular endothelial surface. Protein C is a vitamin K-dependent serine protease synthesized and secreted by the liver as a zymogen which, together with protein S (also synthesized by the liver) and a vascular endothelial protein (thrombomodulin), provides a major naturally occurring anticoagulant system. When thrombin is generated intravascularly it binds to thrombomodulin on the cell surface of the vascular endothelium. This complex activates protein C, which then forms a complex with protein S on the surface of both vascular endothelium and platelets. This protein C–S complex inactivates factor Va and factor VIIIa and plasminogen activator inhibitor. Thus, in addition to abrogating clotting on these surfaces, the protein C pathway stimulates fibrinolysis. The extrinsic pathway inhibitor (EPI) is a low-molecular-weight plasma protein that stops coagulation initiated by the factor VIIa–tissue factor complex.

Other vascular endothelial defences against thrombosis include the secretion of tissue plasminogen activator (promotes fibrinolysis) and the elaboration of prostacyclin, which is a potent vasodilator and inhibitor of platelet adhesion.

The haemostatic response

The haemostatic response, which prevents exsanguination after vascular injury, has two components:

- primary response: vasoconstriction and formation of a platelet plug
- secondary response (haemostasis): formation of a fibrin seal that ensures continued haemostasis until healing of the vascular injury is complete.

Primary response

The factors that initiate and maintain the vasoconstriction are not known, although platelet factors such as thromboxane A_2 [(TXA_2) a potent vasoconstrictor] and substances released from the endothelium are probably involved. Contrary to former belief, serotonin does not mediate the initial vasoconstrictor response.

The formation of the platelet plug involves three interrelated stages: platelet adhesion, the release reaction and platelet aggregation. Platelet adhesion to the vascular endothelium and adjacent perivascular connective tissue (predominantly collagen) is the initial event and appears to be receptor mediated [von Willebrand factor (vWF)]. Congenital absence of this factor (von Willebrand disease) results in a prolonged bleeding time with a normal clotting time. Following adherence to collagen, the platelets change shape from flat discs to spheres with polypoidal projections and actively secrete a number of factors from their α-granules (release reaction). These include ADP, β-thromboglobulin, platelet factor-4, platelet-derived growth factor and vWF. ADP is largely responsible for the platelet aggregation and this is augmented by TXA_2 derived from arachidonic acid (released from platelet membrane phospholipids) via the cyclo-oxygenase pathway. Actual aggregation of platelets to each other requires the presence of fibrinogen that binds to specific surface receptors on several platelets and links them together. Patients with congenital absence of fibrinogen (afibrinogenaemia) or absence of the specific platelet receptors for fibrinogen (Glanzman's thrombasthenia) have prolonged bleeding times owing to defective platelet aggregation. Several of the clotting factors synthesized by megakaryocytes (fibrinogen, factor V, etc.) are stored in platelets, from which they are liberated during the release reaction, thus contributing to the coagulation cascade. In addition, activated platelets provide a procoagulant factor (PF3), which enhances the activation of prothrombin to thrombin by factor Xa.

Secondary response (blood coagulation)

The essence of blood coagulation is a chain reaction or cascade of proenzyme to enzyme conversions, with each enzyme activating the next proenzyme until thrombin acts on fibrinogen to produce fibrin monomer, which then polymerizes to stable fibrin. The various clotting factors are shown in Table 1.2.

By convention, two pathways of blood coagulation are recognized: extrinsic and intrinsic. Both lead to the activation of

Table 1.2 Coagulation factors

Designation	Synonym
I	Fibrinogen
II	Prothrombin
III	Tissue thromboplastin
IV	Calcium ion
V	Proaccelerin, labile factor
VII	Serum prothrombin conversion accelerator
VIII	Antihaemophilic factor
vWF	von Willebrand factor
IX	Christmas factor
X	Stuart–Prower factor
XI	Plasma thromboplastin antecedent
XII	Hageman factor
XIII	Fibrin stabilizing factor
Prekallikrein	Fletcher factor
HMW kininogen	Contact activation factor

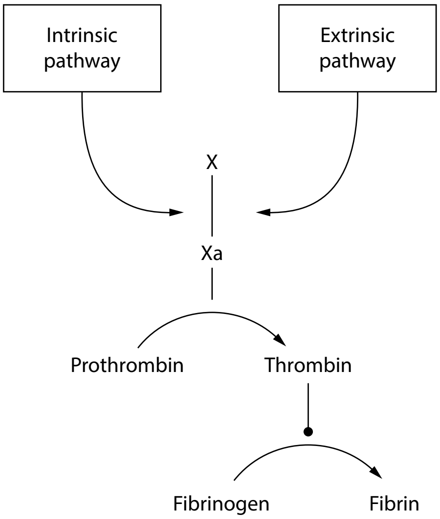


Figure 1.17 The two pathways of blood coagulation: extrinsic and intrinsic. Both lead to the activation of factor X (Xa) and thereafter follow the same cascade reaction.

factor X (Xa), and thereafter follow the same cascade reaction (Figure 1.17). This division is, however, somewhat arbitrary as the two systems converge at other levels (e.g. factor IX). However, the distinction is of practical value as, while the prothrombin time (PT) reflects the activity of both pathways, the partial thromboplastin time measures the activity of the intrinsic system only. The extrinsic system is initiated by exposure of blood to injured tissue via the release of tissue thromboplastin (i.e. tissue factor that is extrinsic to the blood). By contrast, the intrinsic pathway is triggered by contact of Hageman factor with a foreign surface and no extrinsic component is needed for the coagulation process. A simplified system of the blood coagulation cascade is shown in Figure 1.18.

Formation of a stable fibrin clot

Thrombin cleaves four fibrinopeptides from fibrinogen with the formation of fibrin polymer, which polymerizes spontaneously to form fibrin. The stabilization of this to yield the insoluble fibrin requires the action of XIIIa (activated by thrombin) and calcium ions.

Fibrinolysis

The lysis of fibrin by plasmin is an important physiological mechanism that guards against occlusion of blood vessels and is important in tissue repair. Following haemorrhage due to vascular injury, fibrinolysis is temporarily switched off. Premature activation will result in renewed bleeding and this mechanism may be responsible for recurrent bleeding from gastrointestinal ulceration. Abnormal fibrinolysis is associated with severe bleeding disorders and is involved in tumour spread. There are both intrinsic (factor XII, prekallikrein, high-molecular-weight kininogen and prourokinase) and extrinsic (tissue plasminogen activator, urokinase) activators of plasminogen. Inhibition is achieved by the activities of plasminogen activator inhibitor 1 (produced by endothelial cells, hepatocytes and fibroblasts) and α_2-plasmin inhibitor (Figure 1.19).

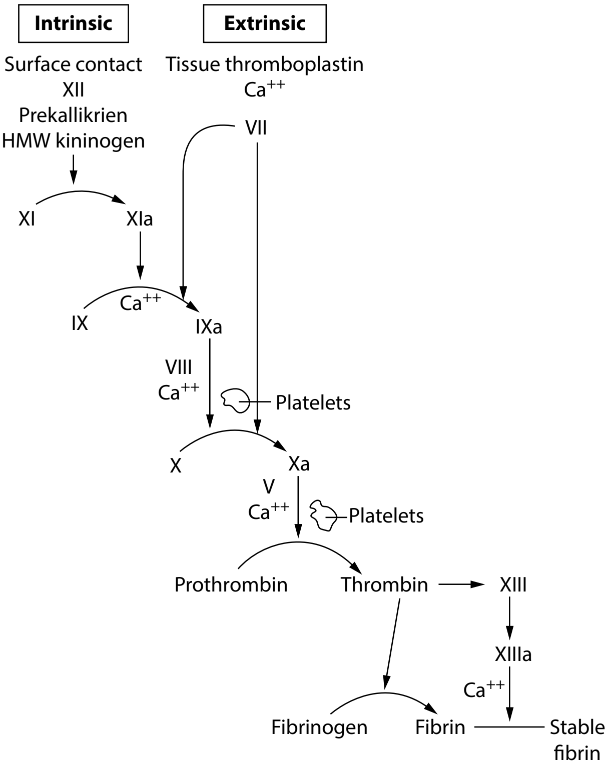


Figure 1.18 Simplified system of the blood coagulation cascade.

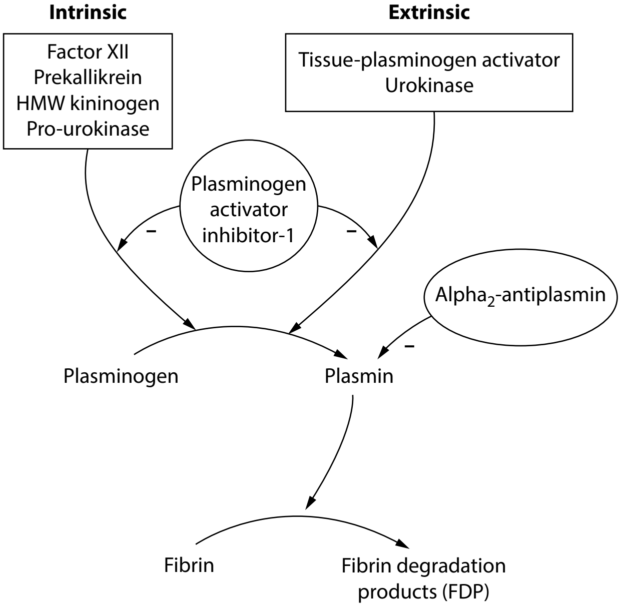


Figure 1.19 The fibrinolytic system.

Therapeutic anticoagulation

In surgical practice, anticoagulation is used:

- in patients with acute intravascular thrombosis (arterial and venous)
- in patients with pulmonary embolism
- as prophylaxis against deep vein thrombosis
- in patients with an implanted intravascular synthetic prosthesis, most commonly heart valves
- patients with atrial fibrillation.

Therapeutically, the anticoagulants used are heparin (administered intravenously), low-molecular-weight heparins (LMWHs) (administered subcutaneously) and warfarin for oral anticoagulation.

Heparin and low-molecular-weight heparins

Heparin consists of a mixture of anionic mucopolysaccharides (glycosaminoglycans): α-L-iduronic acid 2-sulphate; 2-deoxy-2-sulphamino-α-D-glucose 6-sulphate; β-D-glucuronic acid; 2-acetamido-2-deoxy-α-D-glucose; and α-L-iduronic acid. In heparin sodium (used clinically), the acidic protons of the sulphate units are partially replaced by sodium ions. Heparin acts by immediate binding to AT-III. Only about one-third of an administered dose of heparin binds to AT and this fraction is responsible for most of its anticoagulant effect. The dose of heparin is monitored by measurement of the activated partial thromboplastin time (APTT) and its action can be reversed by protamine sulphate (see below). The important attribute of heparin is its immediate anticoagulant effect; because of its potency, it is only administered in hospital practice. Thus, when long-term anticoagulation is needed, patients are started on heparin, but warfarin is commenced at the same time. When this results in the correct prolongation of the PT the therapeutic range is reached (see below), the heparin is discontinued.

Low-molecular-weight heparins

LMWHs possess the following characteristics with respect to heparin:

- reduced antifactor IIa activity relative to antifactor Xa activity
- more favourable benefit–risk ratio
- superior pharmacokinetic properties.

LMWHs were introduced in the 1970s and prepared from standard heparin as derivatives that exhibited a progressively reduced effect on the APTT with decreased molecular size, while still retaining the inhibitory effect on activated factor X (factor Xa).

LMWHs are derived from heparin by chemical or enzymatic depolymerization, yielding fragments approximately one-third the size of heparin. The various LMWHs approved for clinical use are:

- nadroparin calcium (Fraxiparin)
- enoxaparin sodium (Lovenox/Clexane)
- dalteparin (Fragmin)
- ardeparin (Normiflo)
- tinzaparin (Innohep)
- reviparin (Clivarine)
- danaparoid sodium (Orgaran).

LMWHs produce their anticoagulant effect by activating AT-III via a unique pentasaccharide sequence. Laboratory monitoring of LMWH therapy is usually not necessary except in morbid obesity and renal failure, where the dose can be difficult to determine. In these patients, laboratory assay is often needed, i.e. the chromogenic anti-Xa assay. High anti-Xa levels (>0.8 U/mL at steady state) in patients receiving therapeutic doses of LMWH have been associated with bleeding in clinical studies. The optimal time to perform an anti-Xa assay (when indicated) is 4 hours after subcutaneous injection of a weight-adjusted dose of LMWH.

The clinical indications of LMWHs are:

- prevention of venous thrombosis
- unstable angina
- Q-wave acute myocardial infraction
- coronary angioplasty.

Warfarin

Warfarin (coumarin derivative) was originally introduced as a rat poison. It acts as a vitamin K antagonist, which is essential for the synthesis of clotting factors II, VII and IX and anticoagulant proteins C and S. The dose of warfarin needs to be adjusted to the individual patient to achieve the correct prolongation of the PT [international normalized ratio (INR) 2.0–2.5]. This requires several days with frequent monitoring of the PT. The exact dose of warfarin needed is influenced by age, genetic status, medications, diet and certain medical conditions, all of which may alter patient response. Because of these problems in patients requiring long-term oral anticoagulation, new oral anticoagulants have been introduced: (1) dabigatran, which is a direct thrombin inhibitor, (2) rivaroxaban, a direct factor Xa inhibitor, and (3) ximelagatran, also a thrombin inhibitor.

Newer oral anticoagulants

Dabigatran etexilate is a direct thrombin inhibitor. It acts by binding to thrombin, blocking its interaction with substrates, and thereby prevents the thrombin-induced conversion of fibrinogen to fibrin. Several randomized clinical trials with dabigatran etexilate have confirmed that it is equally efficacious as enoxaparin (LMWH) and has a similar safety profile. The National Institute for Health and Clinical Excellence (NICE) in the UK has approved the use of dabigatran as an alternative to LMWH. Three randomized clinical trials have also confirmed that rivaroxaban, which acts by inhibiting factor Xa, is also as efficacious and is as safe as enoxaparin in the prevention of venous thromboembolism in patients undergoing major orthopaedic surgery. Rivaroxaban has also been approved by NICE for the prevention of venous thromboembolism in adults having elective hip or knee replacement surgery. Thus, both dabigatran etexilate and rivaroxaban may well replace warfarin in patients requiring oral anticoagulation, especially as both oral anticoagulants have a rapid onset of action and predictable pharmacokinetics, enabling fixed dosing and avoidance of frequent monitoring of the PT associated with warfarin therapy.

The latest new oral anticoagulant as a potential substitute to warfarin is ximelagatran (Exanta). This acts through its metabolite, melagatran, which inhibits the conversion of fibrinogen to fibrin by thrombin. Two 20 mg tablets of ximelagatran daily provide an anticoagulant effect equivalent to warfarin without any need to monitor the clotting time. Ximelagatran is not metabolized by cytochrome P450 in the liver, and thus its effects are not altered by food or supplements. Furthermore, there is no known interaction of ximelagatran with other pharmaceutical drugs. Although it does carry a risk of bleeding, this is no greater than that of warfarin, i.e. in the 2.0–3.0 INR range. Clinical trials with ximelagatran have been carried out in patients with myocardial infarction, in the prevention of postoperative deep vein thrombosis and in patients with atrial fibrillation – all confirming efficacy and safety.

Protamine sulphate

This is used to neutralize or reverse heparin anticoagulation. Protamine sulphate is a cationic protein derived from fish sperm that binds strongly to heparin in a ratio of approximately 100 U of heparin to 1 mg of protamine. The administration of protamine sulphate carries certain risks of severe adverse reactions, including hypotension and bradycardia, but these can be minimized by slow administration over 1–3 minutes. Allergic reactions can occur, including anaphylaxis (usually after previous exposure to protamine-containing insulin) and in patients with hypersensitivity to fish protein. Alternative treatments to neutralize/reverse heparin in patients with known hypersensitivity to protamine sulphate include hexadimethrine, heparinase (neutralase), extracorporeal heparin removal devices and synthetic protamine substitutes.

Protamine also neutralizes the antithrombin activity of LMWHs, normalizing the APTT and thrombin time. In animal studies, synthetic protamine variants have been highly effective in neutralizing LMWHs (including anti-Xa activity) and are less toxic than protamine.

Surgical immunology

Normally, the immune response is directed against the antigenic characteristics of micro-organisms and, to a much lesser extent, cancer cells. A breakdown in the immune system can take two forms: (1) a lack of appropriate response to a pathogen or (2) a reaction by the immune system against host antigens causing disease or rejection. In order to understand these processes, it is necessary to describe the components of the immune system.

Major histocompatibility complex

The MHC is central to the immunological identity. The MHC consists of a series of antigens expressed on nucleated cells. Thus, red blood cells do not have these antigens but instead have the blood group isoantigens (A, B).

The MHC was first identified on leucocytes, where it is known as the HLA system, the genes for which are located on the short arm of chromosome 6. The MHC gene region encodes for three groups of antigens: class I, class II and class III. The HLA system is made up of class I and II antigens. The function of class III gene product is not clear. There are three class I loci (A, B, C) and three class II loci (DR, DQ, DP) on the chromosome. Thus, this system is highly polymorphic, there being at least 20 alleles in A, 40 in B, eight in C, 16 in DR, three in DQ and six in DP. The inheritance of MHC genes is Mendelian and the expression codominant, so each individual expresses two alleles at each of the loci.

Class I antigens are expressed on all nucleated cells, whereas class II antigens are found on cells involved in the immune response. Both classes are involved in immune recognition, the molecules forming a binding site for processed antigen to be identified by the T-cell receptor. The reason for the diversity of the MHC system is not completely understood, but, like all genetic diversity, it appears to have a survival advantage, e.g. it enables different individuals to mount different responses to invading micro-organisms, i.e. virulent agents are lethal to some but not to other individuals. The MHC is associated with certain disorders, often with an underlying autoimmune basis, e.g. ankylosing spondylitis, which is 20 times more common in individuals possessing HLA B27, and juvenile insulin-dependent diabetes.

In addition to the MHC, there are numerous other tissue antigens collectively referred to as the minor histocompatibility antigens. The loci for these are widely distributed within the genome. They may be the major source of organ allograft rejection in patients matched at the MHC.

Cellular components of the immune system

Phagocytes

Phagocytes are found in the bloodstream (neutrophils and monocytes) and in the tissues as part of the monocyte–macrophage (reticuloendothelial) system as tissue macrophages, Küpffer cells in the liver, microglia in the brain, mesangial cells in the kidneys and dendritic cells in the skin. Phagocytes of the monocyte type are responsible for processing antigens, pathogens and other particles. After alteration of the phagocytosed material by intracellular lysosomes, the phagocytes present the foreign material or antigens to the lymphocytes, which then initiate the immune response. This process is called antigen presentation, and any cell undertaking this function is termed an antigen-presenting cell (APC). The interaction among the antigen, the APC and the lymphocyte leads to the release of chemical messengers called cytokines. In turn, these cytokines enhance the ability of macrophages and other competent cells to recognize and assimilate further antigenic material. Neutrophils do not act as APCs. Instead, they completely degrade the phagocytosed material.

Phagocytosis

The ability of phagocytes to take up pathogenic material is dependent on surface receptors that can recognize activated complement (C3 system) and the Fc fragment of immunoglobulin (see below), both acting as opsonins, i.e. they coat the invading antigen. The density of these receptors on the cell surface determines the level of phagocytosis and the activity of their intracellular lysosomal enzymes.

The processed antigenic fragments become complexed with class II MHC antigens and migrate to the surface of the APC. Here, the combined processed foreign antigen/class II complex becomes exposed for antigen recognition by the immune system. Most antigen recognition requires the combination of foreign antigen with self-MHC for T-cell activation, but there are exceptions. Thus, MHC alloantigens, e.g. tissue antigens of an organ graft, can be recognized more directly by specific T-cells, presumably because these alloantigens resemble the type of modification to self-MHC normally produced by combination with a foreign antigen.

The most active APCs appear to be the tissue macrophages that originate from blood monocytes and continue to recirculate carrying processed antigen from the periphery to the regional lymph nodes, where they initiate the immune response. In addition, the Küpffer cells lining the sinusoids of the liver act as APCs and present processed antigen to the circulating T-cells.

Dendritic cells

Dendritic cells (DCs) are known as the 'pacemakers' of the immune response in view of their key role as the major APCs of foreign peptides and proteins to both T- and B-lymphocytes. DCs are responsible for the induction of T-cell-mediated immunity. T-lymphocytes have surface cell receptors that are able to recognize fragments of antigens (Ags) bound to the MHC (class I and II) on the surface of DCs and other APCs. The MHC class I peptide-binding proteins interact with and stimulate cytotoxic T-lymphocytes, whereas the MHC class II peptide-binding protein stimulates the T-helper cells. Within the APCs, Ags are processed into peptides before being re-expressed on the cell surface linked to MHC proteins. Cytotoxic T-lymphocytes are generated when the re-expressed antigen is bound to MHC class I molecules, which become primed to destroy cells in tissues expressing the Ags.

DCs are derived from the bone marrow and from progenitor cells to form lymphoid DCs (lymphocytic lineage) and the myeloid DCs (myeloid lineage), which include multipotent CD34+ progenitors and peripheral blood mononuclear cells. DC progenitors are released from the bone marrow and circulate through the blood and lymphoid organs awaiting further differentiation and homing signals. The myeloid pathway of differentiation gives rise to DCs that home to peripheral tissues to take up and process exogenous Ags prior to migrating to the secondary lymphoid tissues to present Ags to naive T-cells. The thymic DCs have a different and very specific purpose: their function is to present the self-Ag to developing thymocytes and, hence, subsequently delete autoreactive T-cells.

DCs are present in the peripheral blood (PBDCs), in the interstitial space of all tissues (with the exception of the cornea and central nervous system), in the thymus, lymph nodes and spleen. In the peripheral blood, two subpopulations of PBDCs have been identified, but both contain a specific protein (p55), and are thus detected by monoclonal antibody against p55. The recruitment of DCs in the peripheral tissues is thought to be mediated by the local production of cytokines such as granulocyte–macrophage colony-stimulating factor (GM-CSF) and by systemic signals such as bacterial lipopolysaccharides. DCs migrate to the secondary lymphoid tissues via the afferent lymphatics, but are thought to reach the spleen via the bloodstream. However, the mechanisms of homing to the lymph nodes and spleen are not fully understood, although certain isoforms of the hyaluronic acid receptor (CD44) are thought to be involved. Within the lymph nodes DCs are situated within the T-cell paracortical regions, whereas in the spleen they are located in the marginal zones at the periphery of the periarterial sheaths.

Lymphocytes

Lymphocytes form 20% of the white cell population and are of two kinds: T- and β-cells. The functional classification of lymphocytes depends on the detection of cell-surface markers that are identified using specific monoclonal antibodies. These cell-surface markers are termed clusters of differentiation (CD) and have been defined by international agreement. Some of these markers are specific adhesion molecules, e.g. CD54, which is intercellular adhesion molecule (ICAM) 1.

T-lymphocytes

T-cells differentiate in the thymus (at least in the fetus) and carry a number of important functions:

- regulation of the level of the immune response
- cytotoxic activity against bacteria
- assistance to β-cells in the destruction of virally infected cells
- initiation of cytotoxic activity of other immune effector cells.

All T-cells carry a receptor (Figure 1.20) that recognizes antigens and MHC molecules. This receptor is related to immunoglobulin and is associated with the CD3 molecule that

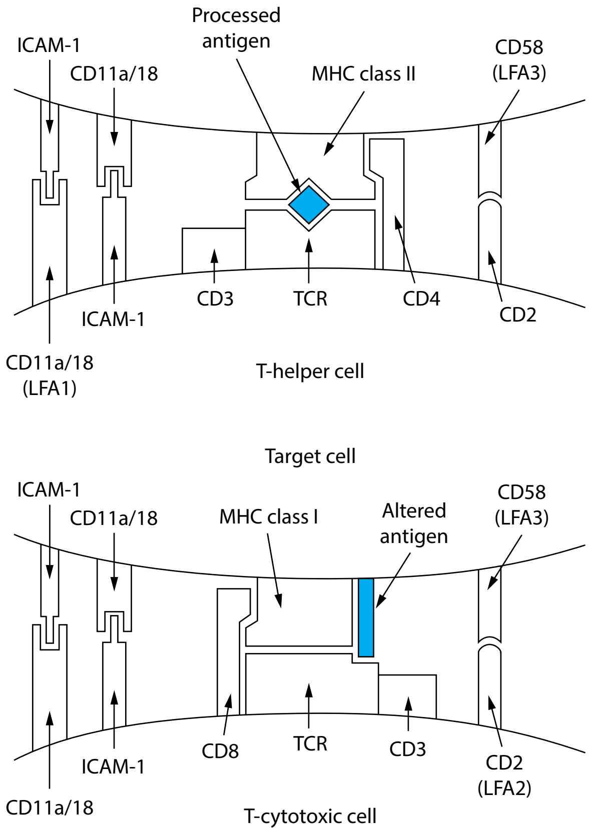


Figure 1.20 Cellular interactions involving the major histocompatibility complex (MHC) and the T-cell receptor (TCR). Processed antigen is presented to the TCR on the T-helper cell by the antigen-presenting cell (APC) together with MHC class II, whereas altered (e.g. viral) cellular antigen is presented to the T-cytotoxic cell with MHC class I. These interactions are also dependent on CD4 and CD8, respectively, and aided by a number of adhesion molecules.

is thus used as a pan-T-cell marker. Both CD3 and the T-cell receptor are members of the immunoglobulin superfamily, as are CD4 and CD6. The T-cell receptor/CD3 complex is essential for the functioning of the immune system. Activated (but not resting) T-cells also carry class II MHC molecules and have receptors for the cytokine IL-2.

There are two main types of T-lymphocyte (helper and suppressor), which are differentiated by their surface markers and function. T-helper cells are characterized by the surface receptor CD4, the ligand of which is part of the MHC class II complex on the APC. Processed antigen is expressed with the MHC class II molecule and is presented to the T-cell receptor/CD3 complex on the T-helper cell, thereby initiating the immune recognition and response. This response involves the T-helper lymphocyte producing several cytokines (particularly IL-2), which are involved in the maturation of cytotoxic T-lymphocytes, B-lymphocytes and various other functions of the immune system.

Cytotoxic T-lymphocytes express the surface receptor CD8, which recognizes class I MHC molecules (expressed in all nucleated cells). When foreign antigen, e.g. from virally infected or tumour cells, is expressed together with MHC class I molecule, this is recognized by the T-cell receptor/CD3 complex on the cytotoxic T-cell. This then releases various proteases that puncture the cell membrane, destroying the abnormal cell.

The T-lymphocytes with suppressor function form an ill-defined group that also expresses the CD8 marker. There is no doubt that suppressor activity exists, but the existence of specific T-suppressor cells remains unconfirmed.

B-lymphocytes

B-lymphocytes develop in the bone marrow and fetal liver. Activated β-cells proliferate and mature under the influence of T-cells. They ultimately differentiate into plasma cells that are characterized by an expanded cytoplasm, a large endoplasmic reticulum and a synthetic system entirely concerned with antibody production. The characteristic surface marker of B-lymphocytes is an immunoglobulin. In addition, β-cells have receptors (C1, C2 and receptor for the Fc fragment of antibody). β-cells produce antibodies (immunoglobulins) in the presence of foreign antigens with the help of antigen-specific T-cells. Immunoglobulin is the β-cell equivalent of the T-cell receptor.

Null cells

Null cells are a distinct group of leucocytes that exhibit neither T-cell nor β-cell characteristics. They carry a mixture of surface markers including CD16 and CD56. In appearance they are large, granular lymphocytes. The presence of receptor for the Fc portion of IgG allows null cells to mediate antibody-dependent cellular cytotoxicity (ADCC). Some cytotoxic T-cells also express the Fc receptor and are thus capable of ADCC (see below).

Natural killer lymphocytes

Natural killer (NK) lymphocytes are able to recognize antigenic determinants on the surface of some tumour and virally infected cells and destroy these without the assistance of antibody, although they are also capable of ADCC. The direct cytotoxic activity is not MHC restricted. NK cells express receptors for γ-interferon and IL-2. They are thought to serve a surveillance role with activity against circulating tumour cells, especially those of lymphoid origin. There is also some evidence that reduced NK activity correlates with tumour dissemination and with postoperative infective complications. Most NK function is undertaken by large granular lymphocytes.

NK cells do not have any cytotoxic activity against solid tumours. However, they can be transformed into lymphokine-activated killer (LAK) cells by incubation with IL-2. LAK cells are cytotoxic against various tumour cell lines. This form of immunotherapy (infusion of ex vivo produced LAK cells from a patient's NK cells) has been demonstrated to show some activity against certain human solid tumours, particularly melanoma and renal carcinoma, but its excessive toxicity has limited its usage and acceptance. Both NK and LAK cells are capable of ADCC.

Cytokines

These are water-soluble, low-molecular-weight peptides that act on or are produced by lymphocytes. They are grouped as follows:

- interleukins: IL-1 to IL-13
- interferons: α-interferon, β-interferon, γ-interferon
- colony-stimulating factors: macrophage colony-stimulating factor (M-CSF), GM-CSF, granulocyte colony stimulating factor (G-CTF).

Cytokines have several functions. In the first instance, some members are involved in the initiation of the inflammatory process. Others promote cell recruitment to the damaged area during the process of repair. Some cytokines are able to act more specifically, e.g. macrophage-stimulating factor enhances the ability of macrophages to ingest micro-organisms. An important function of cytokines is to provide a short-range messenger system of communication between cells of the immune response, hence the name interleukins. Cytokine signalling is extremely complex, and a simplified version is shown in Figure 1.21.

IL-1 is produced by macrophages and, to a lesser extent, by lymphocytes and fibroblasts. It acts as a short-range messenger in combination with antigen to cause T-cells to release IL-2 and probably express the IL-2 receptor (IL-2R). IL-1 also acts as a circulating humoral agent and induces the liver to manufacture acute phase reactive proteins and the hypothalamus to raise the body temperature. It also stimulates β-cells and induces the macrophages to release prostaglandin E_2 and tumour necrosis factor (TNF). IL-1 is responsible for certain symptoms, including arthralgia and myalgia. It also causes weight loss and tissue destruction in certain disorders. The production of IL-1 can be reduced or suppressed by aspirin, indometacin, corticosteroids and antimalarial agents.

IL-2 is produced by T-cells. It acts as a T-cell growth factor causing clonal proliferation in response to an antigen. Most T-cell subsets will proliferate in the presence of IL-2, the extent depending on the density of the IL-2 receptor, the level of expression of which requires restimulation and further exposure to antigen. IL-2 levels are reduced by immunosuppressive agents, e.g. ciclosporin and corticosteroids. The effect of corticosteroids

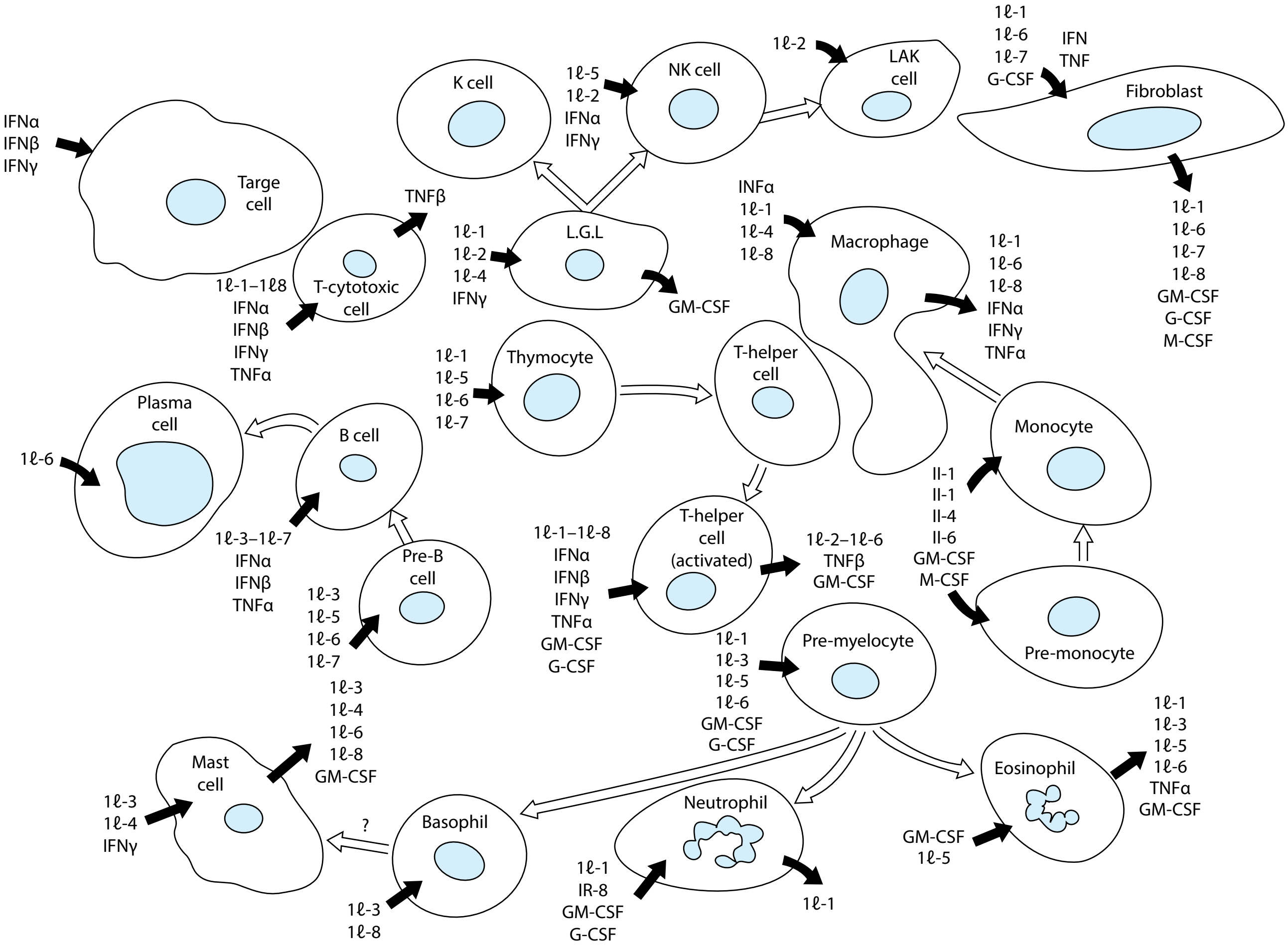


Figure 1.21 Simplified version of cytokine signalling by immunologically active cells. The cytokines produced by and influencing the cells of the immune system are shown. IL, interleukin; IFN, interferon; TNF, tumour necrosis factor; GM-CSF/G-CSF/M-CSF, granulocyte–macrophage, granulocyte and monocyte–macrophage colony-stimulating factors; LGL, large granular lymphocyte; K, killer; NK, natural killer; LAK, lymphokine-activated killer.

is probably secondary to suppression of IL-1 production. IL-2Rs are also found on β-cells, monocytes and null cells.

IL-3 is produced mainly by T-helper cells and mast cells, which produce histamine. It acts as a colony-stimulating factor on all haemopoietic cells. Its presence is obligatory for the development of mast cells.

IL-4 is produced by T-helper cells, for which it acts as an autocrine growth factor. It also stimulates β-cells to proliferate. Activity appears to extend to most haemopoietic cells. IL-4 is of crucial importance in the induction of several immunoglobulins.

IL-5 is produced mainly by T-cells and induces proliferation of both β-cells and eosinophils. It is thus involved in type I allergic reactions. IL-5 acts in synergy with IL-2, in the induction of LAK activity in peripheral blood lymphocytes but is inactive in this respect on its own.

IL-6 is an inflammatory cytokine produced by macrophages, fibroblasts, mast cells and T-cells. It acts on a variety of targets in the immune and haemopoietic systems. Many of the activities thought to be due to IL-1 are actually secondary to the induction of IL-6. It is also a β-cell growth factor, being involved in the production of plasma cells. IL-6 stimulates malignant myeloma cells.

IL-7 is produced by fibroblasts and stromal cells. It stimulates the growth of early lymphoid cells, especially those of the B-lineage.

IL-8 is an inflammatory cytokine produced by stromal cells, endothelium and gastrointestinal mucosal cells. It stimulates neutrophil infiltration.

IL-10 is derived from T-cells and inhibits the production of cytokines from monocytes and macrophages.

Interferons are produced by leucocytes (α-interferon), fibroblasts (β-interferon) and antigen-activated lymphocytes (γ-interferon). The overall role of interferons is to synergize with other cytokines, but α- and β-interferons also enhance class I MHC antigens expression, whereas γ-interferon enhances class II expression. All act as antiviral agents and, together with IL-2, augment NK cell cytotoxicity against certain target tumour cells.

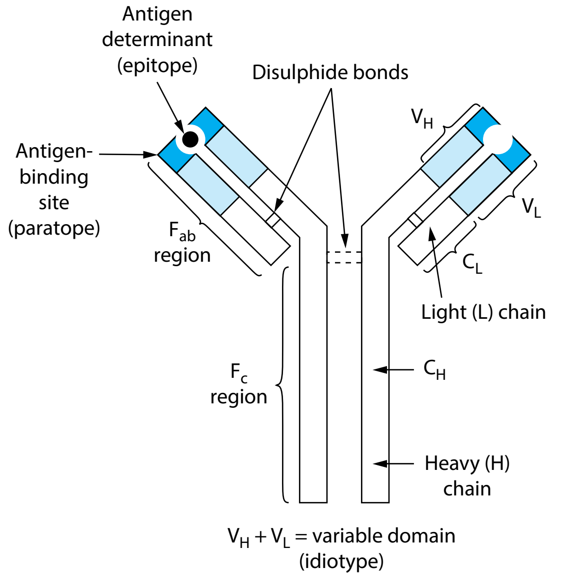
TNF is derived from macrophages and lymphocytes and kills tumour cells in some animal systems but not in humans. TNF exists in two forms, α and b. Cachectin is related to TNF and is possibly the same substance. High levels of TNF are accompanied by wasting of both adipose tissue and muscle and are responsible, together with other cytokines/factors, for wasting syndromes. Thus, prolonged infusion of TNF alone in humans does not result in cachexia. Most of the current interest in TNF relates to its role as one of the mediators of multiple organ failure, although monoclonal antibodies against TNF and endotoxin have proved ineffective against the systemic inflammatory response syndrome (SIRS).

The colony-stimulating factors are manufactured by several cells, including stromal cells, fibroblasts, endothelium and lymphocytes. GM-CSF promotes the growth of granulocyte and monocyte progenitor cells. It has proved useful in aiding the recovery of the bone marrow from cytotoxic chemotherapy and radiotherapy (bone marrow rescue). G-CSF acts in a similar fashion on granulocytes, and M-CSF stimulates cells of the monocyte–macrophage lineage.

Immunoglobulins

Immunoglobulins (antibodies) are classified on a structural basis into types and classes: IgG (four classes), IgA (two classes), IgM, IgD and IgE, and are either membrane bound (surface) or secreted. Each has a non-antigen-binding Fc fragment and an antigen-binding Fab moiety. All immunoglobulins are constructed of two identical light and two identical heavy chains cross-linked by disulphide bonds (Figure 1.22). The N-terminal domains are formed from one heavy and one light chain that serve as the antigen-binding site (paratope) reacting with a specific part of the antigen molecule (epitope). The structure of these domains varies with the different antibodies (variable domain), with the antigenic determinant for each being termed the idiotype. The remaining part of the molecule (Fc fragment and part of the Fab moiety) is constant within classes of immunoglobulin, and this antigenic determinant is referred to as the isotype.

The specificity of an antibody for an antigen is dependent on the variable region, which is coded for by gene segments named V, D and J. Variability is consequent on a process of mutation and genetic rearrangements within these gene segments in the β-cell. By these mechanisms, each individual has an enormous heterogeneity of antibody molecules; in fact, more than could be contained in the genome. Membrane-bound immunoglobulin is found on the surface of the β-cells, where it functions as a cell receptor. This immunoglobulin has the same antigen-binding characteristics as the immunoglobulin that may be produced by cells originating from this lineage.



Antigen determinant (epitope)
Disulphide bonds
Antigen-binding site (paratope)
V_H
V_L
F_{ab} region
C_L
Light (L) chain
F_c region
C_H
Heavy (H) chain
$V_H + V_L$ = variable domain (idiotype)

Figure 1.22 Structure of immunoglobulin of the IgG type. This is formed from two heavy chains and two light chains linked together by disulphide bonds. Both heavy and light chains have constant (C_H and C_L) and variable (V_H and V_L) regions. The C_H/C_L region is termed the isotype and the V_H/V_L region, which contains the binding site (paratope), is termed the idiotype.

Each β-cell can produce only one idiotype. Thus, effective antibody production against a specific antigen results from clonal expansion and maturation of individual β-cells into plasma cells. A particular cell can, however, change the Fc portion, and hence the class of antibody, without changing the characteristics of the antigen-binding site. This is because the genes coding the C and V regions are separate and fragments are only spliced as proteins. This explains why the primary response to an antigen may be mainly IgM, whereas the secondary response is mostly composed of the more avidly binding IgG.

The constant areas of the immunoglobulin molecule are concerned with the interaction between the antibody and the receptor on the cell surface. Many of the effector functions are mediated via the Fc receptors on the cell surface and this permits a variety of functions, e.g. the receptor for the Fc fragment of IgG enables macrophages to phagocytose antigenic particles coated with the antibody, or kill an invading organism (ADCC). Similarly, receptors for IgA located on the deep side of epithelial cells facilitate the transfer of the antibody across the cell to be discharged in the secretions. IgE receptors found on mast cells are involved in the release of inflammatory mediators and lymphokines on contact of these cells with specific antigen, and thus initiate the inflammatory response. Antibodies thus subserve two functions:

1. they bind with antigens via the variable domain
2. they connect with the host's defence effector system through the constant domain.

Immunoglobulin G

This is the most abundant immunoglobulin and it constitutes the major antibody produced during a secondary response. It has the ability to cross the placenta into the fetus and is reinforced by colostral IgG during neonatal life. IgG diffuses readily into the extravascular spaces and thus its predominant role is to neutralize bacterial toxins and bind micro-organisms when it activates complement (classic pathway), thereby inducing chemotaxis and phagocytosis. IgG coating sensitizes the target cells to destruction by killer lymphocytes.

Immunoglobulin A

IgA is found in the secretions of exposed mucosal surfaces, e.g. saliva, nasal mucosa, tears, gastrointestinal tract, bronchial and genitourinary epithelium, as a dimer that is stabilized by a protein synthesized by the epithelial cells (secretory component). The purpose of this mechanism is the prevention of bacterial adherence to the exposed mucosal cells and hence invasion.

Immunoglobulin M

These star-shaped antibodies are also known as macroglobulins because of their high molecular weight. These antibodies induce agglutination and cytolysis. IgM antibodies are the first isotype to appear during the course of an infection and are largely confined to the intravascular compartment. Hence, their primary role is thought to be the containment of bacteraemia.

Immunoglobulin D

IgD is nearly all bound to the surface of some of the circulating lymphocytes. It is probably involved in the control of lymphocyte proliferation and suppression and may be involved in establishing tolerance to certain antigens, although its exact function remains unknown. Some consider it to be an evolutionary relic.

Immunoglobulin E

Only a small subpopulation of plasma cells synthesizes IgE and these are mainly confined to mucosal surfaces. Plasma levels of IgE are normally low but levels rise considerably during parasitic infestations. These antibodies, by coating the parasites, sensitize them to destruction by eosinophils. IgE is also associated with the defence of mucosal surfaces in association with mast cells. Encounter with a foreign antigen induces degranulation of the mast cells with the liberation of vasoactive amines, e.g. histamine, causing increased capillary permeability and exudation of plasma factors into the area. It is the mechanism involved in allergic reactions such as extrinsic asthma and hay fever.

Autoimmune disease

Several mechanisms exist to limit the possibility that the immune mechanism may react and damage the host's tissues. In the first instance, clones of T-cells that express self-recognizing receptors are destroyed in the thymus in early life, a process known as clonal deletion. Second, self-reacting clones are functionally suppressed. This is termed clonal anergy. T-cells are crucial in immune regulation. Thus, it is often possible to detect self-reacting β-cell clones but, in the absence of specifically reacting T-helper cells (deleted in the thymus), they are unable to proliferate. In addition, other suppressor mechanisms specifically involving T-cytotoxic cells appear to limit the host response to self.

Nonetheless, these protective mechanisms may fail. Fortunately, although many individuals appear to suffer a breakdown in tolerance towards their own antigenic make-up, only a few develop autoimmune disease. Individuals who develop autoimmune disease show an alteration in the T-helper/cytotoxic ratio in their peripheral blood because the suppressor mechanism has failed. Sometimes there are alterations in the target organs, e.g. thyroid and pancreas, when cells from these organs express class II MHC antigens that are not found in normal individuals. This abnormal expression can be induced by infection, normal physiological triggering, e.g. lactation, or by the action of interferons and TNF-α. In this situation, class II MHC expression is invariably associated with lymphocytic infiltration, which explains the histological appearance of Hashimoto's thyroiditis (see Chapter 18).

Autoimmune disorders are classified as

- organ specific: primary thyrotoxicosis (Graves disease), Hashimoto's thyroiditis, Addison disease, type 1 diabetes mellitus and pernicious anaemia (megaloblastic anaemia, atrophic gastritis and achlorhydria)
- non-organ specific: rheumatoid arthritis, systemic lupus erythematosus, chronic active hepatitis and primary biliary cirrhosis.

Autoantibodies can be detected in the serum of patients with these disorders, e.g. IgG autoantibody to thyroid-stimulating hormone in Graves disease, IgM autoantibody (rheumatoid factor) to the patient's own IgG in rheumatoid arthritis, and antibodies to double-stranded DNA (antinuclear factor) in systemic lupus erythematosus.

The organ damage is induced by the cytotoxicity of macrophages, NK cells and β-cells primed with the respective autoantibody, as well as cytotoxic T-lymphocytes.

The inflammatory response

The inflammatory response can be acute, subacute or chronic. The clinical histology (especially the cellular response) varies accordingly, with neutrophils predominating in acute inflammations and lymphocytes and macrophages in chronic and subacute inflammations. Some conditions are primarily chronic in nature, while others begin as acute but progress to a chronic stage, especially if treatment is suboptimal. Alternatively, acute episodes of inflammation may subside for a time (remissions) with reactivation of the inflammatory process and renewed symptoms and signs (relapses) during the natural history of the disease. This is exemplified by inflammatory bowel disease.

Although in clinical practice infection by micro-organisms is the most important cause of inflammation, this may be caused by non-infectious putative agents: chemical agents, antigen–antibody complexes, trauma, etc. Indeed, traumatic inflammation is an essential part of the healing/repair process and accounts for the changes of the 'lag phase' of healing described earlier in this module.

Acute inflammatory response

The well-recognized features of acute inflammation are pain, swelling, erythema and heat. All of these manifestations are due to the release of chemical mediators. The response varies with the nature of the putative agent. The inflammatory response related to infection is best considered as a mechanism that allows the rapid accumulation of immunologically competent cells and phagocytes to destroy the invading micro-organisms. In this respect, the host's response depends on a number of factors, including the type of organism and the infecting dose, previous exposure and the site of invasion.

Complement fragments C3a, C5a and bacterial endotoxin activate neutrophils, rendering them adherent to the vascular endothelium. Endotoxin also stimulates macrophages to produce and release IL-1 and TNF, itself a potent neutrophil activator. Both cytokines increase the adhesiveness of endothelium for neutrophils.

Various products of arachidonic acid metabolism, including platelet-activating factor, leukotriene B4, the peptidoleukotrienes C4, D4 and E4, and TXA_2, are also involved in the inflammatory response.

The first stage in the acute inflammatory response consists of the migration of neutrophils through the vascular endothelium to the inflammatory source. This involves adherence to the endothelium. The activated neutrophils leave the central column of blood cells of the flowing blood and tumble on the activated endothelial lining before they stick (a phenomenon known as rolling). The binding itself is mediated by two adhesion molecules, L-selectin (leucocyte adhesion molecule 1) and E-selectin (endothelium leucocyte adhesion molecule 1). This is followed by the binding of other molecules known as b_2-integrins on the neutrophils with ICAM-1 and ICAM-2 on the vascular endothelium.

The fully activated extravasated neutrophils produce elastase and free radicals or species, e.g. superoxide ions and hydrogen peroxide, designed to destroy the invader but which can also damage the host's own tissues. This mechanism, with the production of highly destructive reactive oxygen species, can be disadvantageous to the host, as in the ischaemia–reperfusion injury and in certain specific infections such as tuberculosis and leprosy. It may also be involved in SIRS.

Immunocompromised patients

Patients with immune deficiency are unable to mount an appropriate defence against invading organisms and are thus liable to serious infections; moreover, they may show minimal signs in the presence of severe disease. The risk is determined by the severity of the immune deficiency and, in the clinical context, this is more important than the underlying cause. Low-risk patients are more susceptible to infections by organisms that are known to be pathogenic to immunologically competent individuals. Patients with moderate immune deficiency are additionally susceptible to infections by organisms that are normally considered to be harmless commensals. Malnourished patients, patients on steroid therapy and organ transplant recipients belong in this moderate-risk group. In patients within the high-risk group (severely immunocompromised), the risk of fatal infections with both pathogenic and commensal organisms is so high that they may require isolation and the use of prophylactic antibiotics.

Minor degrees of immune deficiency are very common in surgical practice and are frequently overlooked. Some of the factors often encountered in surgical practice include:

- age: neonates, infants and advanced age
- metabolic factors: chronic renal and liver disease
- drugs: antibiotics, non-steroidal anti-inflammatory drugs, steroids, gastric antisecretory drugs, tobacco
- specific disorders: diabetes mellitus, advanced malignancy, collagen disorders, rheumatoid arthritis, cystic fibrosis
- malnutrition
- surgical procedures: intravenous cannulas, bladder catheterization, endotracheal intubation, nasogastric intubation, abdominal drains.

Aside from cystic fibrosis, congenital disorders causing immune deficiency are rare and include congenital neutropenia, deficiencies in antibody production and cell-mediated immunity.

Serious immunosuppression

The causes include drug-induced bone marrow failure from cytotoxic chemotherapy, total body irradiation and graft-versus-host disease and AIDS. A severe degree of immunosuppression results from total body irradiation (1000 Gy), normally used to treat residual leukaemia. This therapy destroys most immunocompetent lymphocytes and there is a decrease in the granulocytes and monocytes, leading to impaired reticuloendothelial activity and antibody formation. Furthermore, radiation also causes chromosomal damage so that the immunoglobulin structure may be altered and less effective. Severe infections *per se* can cause a fall in the neutrophil count in addition to a defect in neutrophil function, detected as toxic granulation in the circulating neutrophils.

The risk of infection is proportional to the level and duration of the neutropenia and is compounded by the anaemia and thrombocytopenia that accompany bone marrow failure. Organisms that commonly cause infections in these patients include pyogenic and endogenous enteric bacteria. Patients are also at risk from exogenous enteric pathogens, e.g. *Pseudomonas, Klebsiella* spp., *Serratia, Actinobacter* and endogenous *Candida* spp. and airborne *Aspergillus*.

The management of these patients is primarily directed to the prevention of infection until the severely immunocompromised state recovers or is reversed by bone marrow rescue or transplantation. Measures used to achieve this objective include strict personal hygiene, use of prophylactic antibiotics including gut decontamination, protective isolation in rooms with a laminar flow system and controls over food and visitors.

Mucosal barriers

The skin clearly has a major role in protection as a physical barrier against invading micro-organisms, but it is also immunologically active, as exemplified by its resident large numbers of antigen-presenting dendritic cells. The protective skin barrier is breached by extensive trauma, including burns, in normal individuals. Immunocompromised individuals are prone to significant infections even after mild skin injuries. Pressure sores (decubitus ulcers) are important sources of infection, common organisms being *Staphylococcus aureus* and *Staphylococcus epidermidis*. In compromised patients, colonization with *Pseudomonas* and *Candida* is common in moist skin areas, and enteric organisms are prevalent on the perianal skin.

The gastrointestinal tract also acts as a barrier and harbours immunologically competent cells. The mucosa of the normal gastrointestinal tract is protected by both mucin and secreted IgA, which prevent the adherence of pathogens. Some pathogens are capable of destroying the mucin protection and, in the event of mucin deficiency, certain normal enteric organisms are capable of attacking the intestinal mucosa. In a similar fashion, gastric acid acts as a strong barrier to infection, but in the event of gastric haemorrhage, gastric neoplasm and the use of gastric antisecretory drugs, e.g. H_2-receptor blockers and proton pump inhibitors, this acid barrier is lost or its efficacy reduced. Normally, commensal organisms with the gastrointestinal tract (the gut microflora) adhere to the mucosa. Invading micro-organisms have to compete for binding sites and available nutrients. This phenomenon is termed colonization resistance. If a patient is prescribed a broad-spectrum antibiotic, especially if this is poorly absorbed, the commensal gut microflora reduces by a factor of 10^3–10^5. This allows an organism that is resistant to the antibiotic used to become established and cause a clinical infection, e.g. *Clostridium difficile* pseudomembranous colitis. Even minor trauma to the mucosal surface, e.g. following endoscopy, can prevent adherence of commensal organisms.

Aside from immune mechanisms, the integrity of the gastrointestinal barrier is dependent on physical continuity of the mucosal lining, adequate blood supply and luminal supply of calories and specific amino acids such as glutamine. There is also evidence that the permeability is increased in severe cholestatic jaundice. The pathological significance of lost or impaired barrier function is the translocation of pathogenic Gram-negative bacteria and/or endotoxins into the portal blood stream. If the Küpffer cell barrier function is overwhelmed, systemic bacteraemia and endotoxinaemia ensue. Thus, significant translocation of bacteria and endotoxin may be encountered:

- after major trauma including major surgery
- in severe sepsis
- in patients on prolonged parenteral feeding
- in patients with intestinal ischaemia
- in patients with severe cholestatic jaundice and liver failure.

A similar pathological mechanism has been implicated in SIRS, which results in multiorgan failure. The widespread damage to the endothelium is now thought to be the result of activation/trafficking of systemic neutrophils. Previously, endotoxin/TNF was held to be primarily responsible, but other substances including cell wall aminophospholipids may be involved in the widespread activation and endothelial damage. In an attempt to reduce this usually fatal condition, selective decontamination of the digestive tract (SDD) has been advocated for critically ill patients within intensive care units. This involves the application of antibiotic and antifungal pastes to the oral cavity and pharynx and oral administration of non-absorbable antibiotics. At least one trial has shown benefit in patients on ventilatory support, although the practice of SDD is still not in widespread usage.

Colonization resistance is also important in the respiratory tract, but only in the upper respiratory passages. In normal individuals, inhaled micro-organisms are rapidly cleared by alveolar macrophages, although this function is impaired in patients with even mild changes in immune function. When an endotracheal tube has been in place for a few days, there is severe mucosal damage, allowing colonization with enterobacteria. Flow of urine is the major factor preventing infection in the urinary tract.

Trauma

Trauma, including surgery, is associated with a multifactorial non-specific immune depression that involves the cell-mediated rather than the humoral response. This is demonstrated by reduced, delayed-type sensitivity on skin testing to recall antigens, e.g. *Candida* and tuberculin. Changes in lymphocyte function that have been documented after trauma and surgery include a decrease in the CD4/CD8 ratio, decreased IL-2 production from T-lymphocytes (CD4+ T-helper cells), reduced NK activity and reduced generation of LAK cells from peripheral lymphocytes. Monocytes are also affected, with the most notable change being reduced class II MHC expression. Indeed, persistent depression of class II MHC expression is associated with a poor prognosis after major trauma. Perioperative blood transfusion adds to the immune depression after surgery and has been shown to be associated with an increased incidence of postoperative infection. Recombinant cytokines and other immunological agents may have a place in patients undergoing major surgery for cancer and in patients sustaining major trauma. Thus, low-dose IL-2 prevents the fall in NK activity and LAK cell generation after surgery. However the clinical efficacy of these agents in these situations remains to be confirmed by prospective studies.

Nutrition

Severe malnutrition has profound effects on the immune system but these are largely confined to the cell-mediated response. All of the commonly measured tests of cell-mediated immunity are depressed. Thus, there is a generalized lymphopenia with marked depletion of the CD4+ T-helper cells, impaired IL-2 production, reduced lymphocyte proliferative response to mitogens, reduced NK activity and generalized atrophy of the lymphoid organs. As with severe trauma, the delayed-type sensitivity to recall antigens is severely impaired and in extreme cases the patient fails to react (anergy). There are no significant changes in the humoral immunity, and the β-cell and immunoglobulin production is not impaired.

Surgical pathology

Surgical pathology involves the macroscopic and microscopic examination by an expert pathologist of surgical specimens removed by surgical operations in the treatment of patients with (1) diagnosed or (2) undiagnosed disease. In the latter instance, the histological examination by the pathologist provides the definitive diagnosis. In cancer, whre the diagnosis is established preoperatively, the purpose of the pathological examination is to:

- establish that the entire tumour has been removed locally by histology of the tumour margins (tumour-free or involved distal, proximal and circumferential margins)
- provide confirmation of the diagnosis and grading of the cancer
- provide a quality control of the locoregional lymphadenectomy performed (adequate or inadequate lymph node harvest)
- determine involvement or otherwise of the regional lymph nodes.

All these affect the subsequent management of the patient and influence the ultimate prognosis. In addition, on the basis of the detailed histology, the surgeon is able to determine that the resection undergone by the patient is:

- potentially curative
- potentially non-curative
- definitively non-curative, which includes debulking procedures.

It follows (without saying) that all surgical specimens and biopsies must be subjected to surgical pathological examination, and it is the duty of the surgeon to ensure that the specimen reaches the pathology laboratory fixed or fresh (depending on the individual need) in a pristine state and that the specimen is clearly labelled in accordance with patient's details.

Tumour margins

The tumour margins relate to the edges of the surgical specimen containing the tumour and determine whether the resection has been locally removed completely, i.e. that the resection has been through normal tissue beyond the limits of the tumour. The extent of the tumour margin will vary with the biological aggressiveness of the tumour and thus will vary from 3–5 mm in well-differentiated tumours to 10–15 mm or more for dedifferentiated (anaplastic) lesions. Although emphasis is placed on proximal and distal tumour margins, it must be remembered that a surgical specimen has a three-dimensional configuration, and a more complete assessment is obtained by including the circumferential tumour margin. This assumes great importance in total mesorectal resection for rectal cancer.

The tumour margins can be examined by the surgical pathologist in different ways: (1) perpendicular section taken at a right angle to the edge of the specimen and or (2) by parallel sectioning.

The perpendicular section enables the important distinction between a lesion that truly extends to the margin and a lesion that comes close but does not involve the margin.

In contrast, a shave section made parallel to the edge of the specimen obtains a relatively large surface area of the margin and is used in determining margins in tumours arising in hollow viscera. However, it does not effectively demonstrate the relationship between the edge of the tumour and the resection margin, i.e. it is unable to identify that the tumour extends to but does not actually involve the margin. Thus, parallel sections are used when macroscopically the margin is large and obviously free of tumour or when dealing with small luminal organs.

Node sampling

The primary purpose of node sampling is to determine the histological lymph node status of the solid cancer. An equally important but often overlooked role is for the pathologist to provide an indication of the oncological appropriateness of a curative operation. Thus, for example, the specimen from a D2 resection for gastric cancer (dissection includes the second tier of lymph nodes draining the tumour) should contain at least 40–50 regional nodes provided the surgical pathologist has harvested the nodes by the appropriate technique as there are many pitfalls, e.g. the lymph nodes may be small and embedded in fat. Many pathologists prefer to search in the fresh specimen, relying on palpation that is not possible if the tissues have been hardened by fixation. Some pathologists prefer to immerse specimens in certain clearing agents to remove excess fat. The nodes collected during the harvest are labelled in appropriate anatomical groups whenever possible. Every lymph node obtained is submitted for microscopic examination after staining with haematoxylin and eosin (H&E). Lymph nodes larger than 5 mm are serially sectioned at 2–3 mm intervals.

Sentinel node biopsies

Aside from routine histological examination of serial sections, special immunohistochemical stains and reverse transcriptase polymerase chain reaction are used to detect small clumps of tumour cells.

Tumour grade

The grade of a tumour is used by the pathologist to estimate the biological aggressiveness in terms of potential for growth, invasion and metastasis of a malignant tumour in a given patient. It is a reflection of the extent of differentiation (how closely the tumour cells approximate to the normal) or dedifferentiation (anaplasia) of the cells constituting the tumour based on the microscopic appearances, i.e. degree of atypia, ploidy, mitotic index, etc., using grades (G1, most differentiated; G4, least differentiated). For some tumours, the tumour grade is based on counting the number of mitoses per 10–50 high-power fields (mitotic index). Grade 1 is a well-differentiated tumour (low-grade malignancy) and grade 2 is moderately differentiated (intermediate grade). The worst prognosis is seen with grade 3 (poorly differentiated, high grade) and especially with grade 4 (totally undifferentiated/anaplastic tumour, high grade). In some tumours, the tumour grade cannot be assessed, in which case the designation GX is given. Individual cancers have their own recognized grading system, e.g. Bloom–Richardson grading for breast cancer, Gleason grading for prostate cancer, etc.

Types of resections

The detailed pathological examination will also modify (usually downgrade) an operation. Thus, a surgeon may consider that he has performed *a potentially curative D2 resection* with no macroscopically visible residual tumour (R0 resection). If the pathological examination confirms clear tumour margins and no histological node involvement (provided the node harvest was of the order expected for the resection), the resection is classed as *definitively curative*. In contrast, even if a few nodes are involved on histological examination, the operation is downgraded to *potentially non-curative*. Positive tumour margins relegate the resection to *definitively non-curative*.

Pathological techniques

Perhaps the most important requirement is the establishment of good communication between the surgeon and the pathologist in terms of the provision of relevant clinical information, correctly obtained biopsy specimens and the need or otherwise for their fixation prior to dispatch to the pathology department. In some situations special laboratory techniques are required in the handling of specimens. These necessitate prior consultation with the pathologist to ensure that the laboratory is set up for the correct handling of the specimen when it arrives. The pathologist can only report on the material received. This may not always be representative of the disease process. In this respect, the surgeon must have the necessary experience and knowledge of surgical pathology to be able to obtain the appropriate biopsy of the lesion in any given situation.

Tissue harvest: biopsy techniques

Tissue diagnosis by the examination of biopsy material, histology, is the final arbiter in establishing the nature of the disease process. The alternative, which is applicable to certain situations and may indeed complement histology, is cytology. This attempts to characterize the cell yield obtained as normal, hyperplastic or neoplastic. The accuracy of both techniques depends on three factors: the skill in obtaining the right representative material, adequate processing and staining of the specimen, and the interpretative skill of the histopathologist or cytologist.

Biopsy

By definition, this involves the removal of a piece of tissue for the examination of the histological architecture and cellular details. It can be open (obtained during surgery), endoscopic or closed (percutaneous). The last two are always partial, whereas open biopsy can be incisional or excisional (total).

Percutaneous closed biopsy

This involves the targeted removal of a needle core of tissue from the suspected site. When this is deeply situated it can be performed 'blind', as in percutaneous liver biopsy, or under visual control (guided biopsy) by radiographic techniques (fluoroscopy, CT), ultrasound or laparoscopy. The type of biopsy needle varies with the nature of the tissue involved, e.g. Abrahams needle for pleural biopsy or Menghini needle for liver biopsy, although the most commonly used implement for soft-tissue biopsy (including liver) nowadays is the Tru-cut needle. Automatic, spring-loaded modifications of this needle, e.g. Biopty device (Figure 1.23), are used for guided needle-core biopsy. With the advent of Tru-cut needle systems, the use of high-speed drill biopsy has become largely confined to bone.

Open biopsy

Open biopsy can be incisional or excisional. Incisional biopsy entails the removal of a wedge of the lesion with adjacent normal margin, whereas in excisional biopsy the entire lesion is excised with a surrounding rim of normal tissue. In practice, all small lesions should be excised *in toto*. Incisional biopsy is used for large lesions prior to treatment.

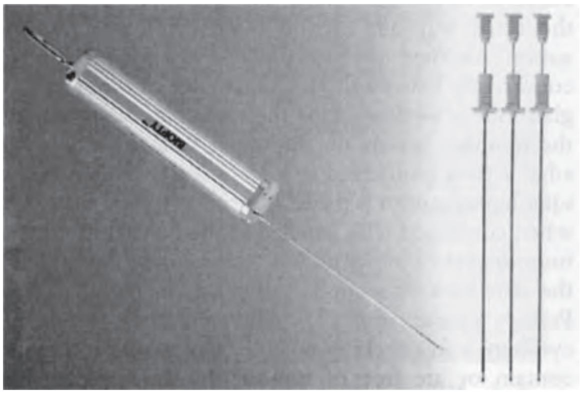


Figure 1.23 Biopty device: a mechanized form of Tru-cut needle-core biopsy.

Endoscopic biopsy

This is now the most commonly performed tissue-sampling procedure. The technique is used in association with both rigid and flexible endoscopy with a variety of biopsy forceps, which have cutting cup-shaped jaws. As the specimens obtained are small, interpretation may be difficult and multiple biopsies are desirable. Endoscopic biopsy is often accompanied by brush cytology, especially for upper gastrointestinal tract lesions, and the combined use of the two techniques enhances the overall diagnostic yield. For inaccessible regions such as the small intestine, special techniques are used. This involves the swallowing of a suction biopsy capsule on a lead (e.g. Crosby capsule). When the desired location, as determined by radiographs, is reached, suction is applied and the cutting knife, situated at the outlet of the capsule, is activated thereby enclosing the desired portion of tissue, after which the capsule is retrieved (Figure 1.24).

Rules governing tissue biopsy

The objective is the harvest of a tissue sample with its tissue architecture preserved and that is representative of the underlying pathological process. This may, on occasion, prove problematic and there are a number of practical difficulties with the various techniques that may limit the usefulness of the histological interpretation. In general, however, certain guidelines may be followed:

- The larger the lesion, the greater the number of biopsies needed for adequate evaluation: multiple biopsies overcome the problem of heterogeneity.
- Biopsies should not be taken from the central crater of ulcerated lesions because this region usually consists largely of necrotic tissue. A wedge peripheral slice that includes the adjacent normal margin is likely to be more informative.
- Whenever safe, the biopsy should include the whole thickness of the lesion so that the pathologist can establish the relationship with the surrounding tissue. This is especially important in providing information on the depth of the lesion: used in the staging of some tumours, e.g. malignant melanoma.
- Large, deeply situated masses are often surrounded by a pronounced peripheral tissue reaction that can form a thick layer simulating

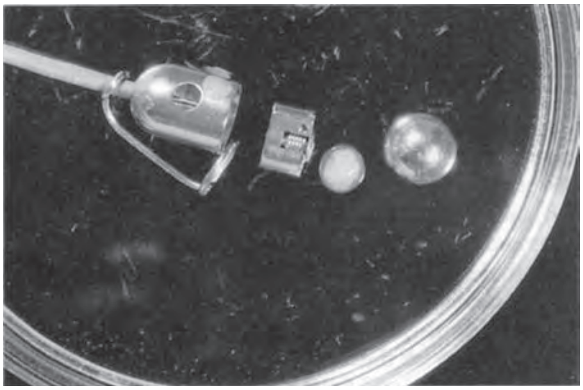


Figure 1.24 Opened Crosby capsule with a jejunal mucosal biopsy specimen in a Petri dish.

a 'capsule'. It is important for the surgeon to ensure that the biopsy wedge transgresses the capsule to include a representative portion of the lesion itself.

- When electrocautery is used to obtain the wedge biopsy, the surgeon should use a cutting current and the biopsy should be larger to ensure that there is sufficient tissue within the wedge that is suitable for histological interpretation.
- The tissue should be handled gently with minimal grasping of the wedge to avoid crushing damage and distortion of the histological features.
- Vascular lesions should be biopsied with extreme care and biopsy should not be attempted unless the surgeon is able to deal with any resulting haemorrhage. These lesions should be excised *in toto* if considered necessary on clinical or pathological grounds.
- Special considerations apply with regard to the biopsy of lymph nodes. This procedure should be considered only after a detailed examination of the patient and the non-invasive appropriate tests have been carried out. The following must precede all lymph node biopsies: search for a primary tumour, exclusion of infectious disease, documentation of the generalized or localized nature of the lymphadenopathy and, in the case of cervical lymphadenopathy, a full ear, nose and throat examination. Whenever possible, lymph nodes should be biopsied intact with the removal of a single node or a group of nodes. When this is not possible because of fixation and matting, a good wedge that includes the capsule and surrounding stroma of the mass is obtained. All lymph node biopsies should be sent fresh to the pathology department, as they often require special immunohistochemical staining and electron microscopy.
- Orientation is essential when the surgeon requires information on resection margins and staging. This has to be performed by the surgeon at the time of excision using easily identifiable markers (e.g. sutures of different materials or lengths), as it may be impossible for the pathologist to orientate the specimen after it has been excised.
- Special considerations apply to liver tumours.

Cytology

In terms of the techniques used, cytology may be exfoliative, brush, imprint or aspiration. The last is obtained by aspiration of fluid or solid masses using a fine-gauge needle and, for this reason, it is commonly referred to as fine needle aspiration cytology or FNA. It is used most commonly by surgeons for the diagnosis of fluid collections (ascites, pleural effusion, etc.), breast lumps and other subcutaneous lesions, thyroid swellings and prostatic enlargement, although it is applicable to a wide variety of conditions and is used routinely by chest physicians for the diagnosis of pulmonary lesions. For breast lesions, a special aspiration gun is often used (Figure 1.25). This enables the surgeon to fix the lump with the left hand while the lesion is aspirated in several directions by the right hand. Deep-seated lesions within the abdomen (e.g. hilar tumours and renal lesions) can also be subjected to FNA under CT or ultrasound control.

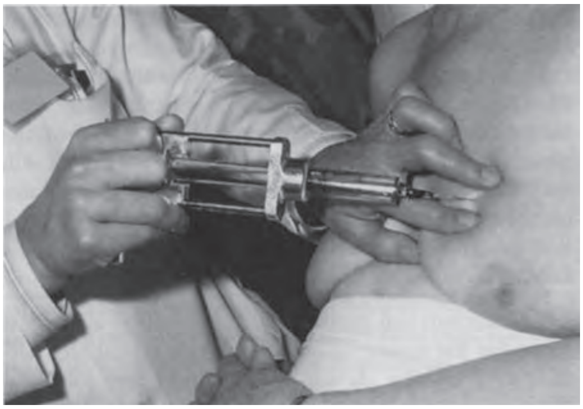


Figure 1.25 Aspiration gun used for fine needle aspiration cytology. It allows the surgeon to fix the lump with the left hand while aspirating the lump in several directions with the right hand.

Brush cytology is employed in endoscopic work. A special brush, introduced through the operating channel of the endoscope, is used and brushings are obtained from the entire surface of the lesion. Some of the superficial cells of the lesions are caught among the bristles and are then transferred to a glass slide. In China, 'blind' brush cytology is used for screening for oesophageal cancer. The patient swallows a small abrasive balloon tied to a piece of string. Once in the stomach, this is pulled up and slide smears are prepared.

Although valuable, imprint cytology is infrequently used by surgeons. One technique entails the application of a sterile glass microscope slide to the cut surface of the tissue, e.g. lymph node. The surface cells become adherent to the glass slide, which is then processed in the usual way and subjected to cytological examination. Another method, scrape cytology, produces a consistently better cell yield. The edge of one end of a glass slide is used to scrape the surface of the tissue and the resulting 'scoop' on the upper surface of the glass edge is then transferred to a second glass slide before a squash preparation is made. Imprint cytology, especially when combined with labelled monoclonal antibody to tumour surface antigens, is a very sensitive method for the detection of secondary deposits in lymph nodes. Perhaps a more surgically useful application of imprint cytology is in checking whether the resection margins contain or are free of tumour. In this respect, it is superior to frozen section, but necessitates the availability of immediate reporting by a skilled cytologist.

Irrespective of the technique used, cytology smears are made on glass slides (both fresh and fixed in carbowax) for staining prior to examination by the cytologist. The stains most commonly used are the Giemsa and the Papanicolaou, but it is important to stress that many of the special stains used in histopathology can be used for the evaluation of cytological material, e.g. stains for glycogen, melanin and mucin. Indeed, it is now possible to determine the oestrogen receptor status of a breast cancer by immunoperoxidase staining of the fine needle aspirate using special antibodies.

Cytological diagnosis is based on the cellular characteristics (nuclear size, chromatin architecture, nucleoli, etc.) and the degree of cohesion of the cells (reduced in cancer aspirates) (Figure 1.26). There are several disadvantages to the existing practice whereby the surgeon obtains the fine needle aspirate of a solid lump and then sends the smears to the cytological department for examination. The most important of these is the high incidence of unsatisfactory samples (smears containing stroma only). This necessitates a repeat of the procedure on a separate occasion. There are several ways in which this problem can be overcome. In some countries such as Sweden, the cytologist takes the fine needle aspirate. In the UK, the preferred option is immediate reporting by the cytologist. This entails the provision of a small laboratory near the clinic. The surgeon hands the fine needle aspirate to the technician, who prepares the smears and stains one with a special haematology stain, 'Diff Quik'. This enables the cytologist to give a definite answer within 5 minutes. Moreover, if the specimen obtained is unsatisfactory, the procedure is repeated immediately.

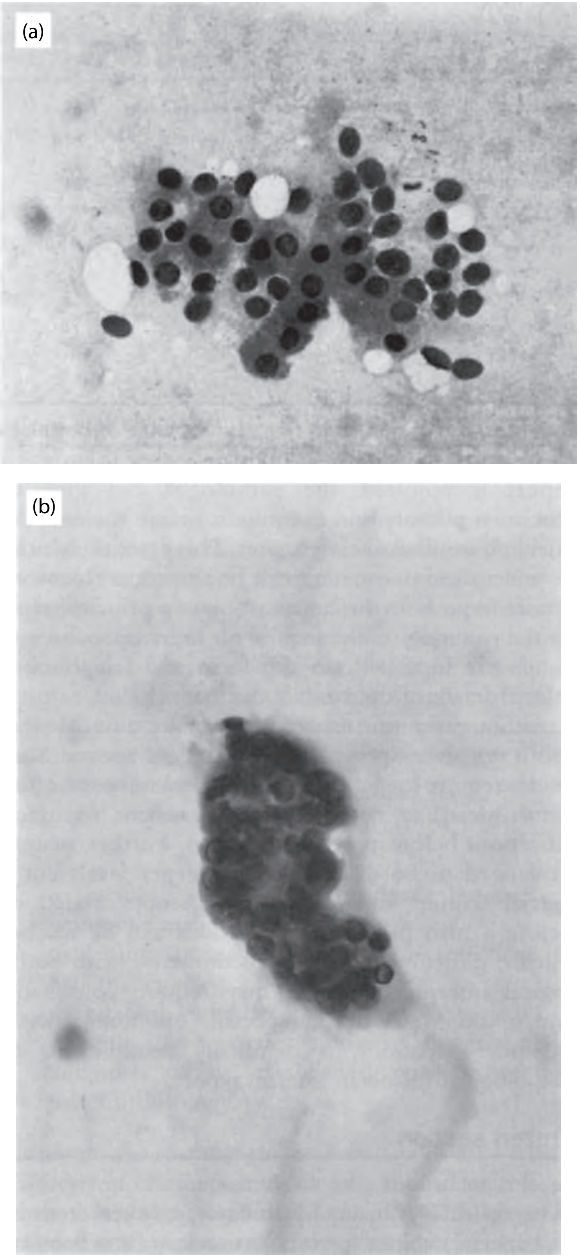


Figure 1.26 (a) Benign cytological aspirate from a fibroadenoma. (b) Malignant aspirate from an invasive breast cancer.

Basic handling of pathological specimens

Fixation

The first consideration relates to fixation of the tissue. Although most specimens should be fixed as soon as possible after removal, there are certain situations in which maximum information is better obtained by the examination of unfixed (fresh) tissue. The usual fixative is 10% buffered formalin and this is suitable for most specimens. However, certain tissues are better fixed in alternative solutions, e.g. Bouin's solution for testicular biopsies and peripheral nerves, chromate solutions for chromaffinomas and glutaraldehyde for tissues to be submitted for electron microscopy. The most important practical consideration relating to the fixation of tissue is the use of a suitably sized container that can accommodate at least 10 times the volume of fixative relative to the specimen. Furthermore, the specimen must be totally immersed in the fixative and the pot accurately labelled.

Certain tissue specimens are best sent fresh to the pathology department for a variety of reasons, e.g. preparation of dabs or smears, culture, enzyme histochemistry and electron microscopy. There must be minimum delay between obtaining the tissue and its arrival at the laboratory and the staff there must be given prior warning and subsequently informed when the specimen is on its way.

Request form

Every specimen must be labelled correctly and accompanied by a completed and signed request form. The information on the request form must correspond to that on the label of the specimen jar. The form must contain details of the following: patient identification, unit and consultant in charge, nature of specimen, date, previous histopathology report numbers (if any), clinical features, operative findings and presumptive diagnosis. If the specimen is considered to be infectious, this should be clearly stated on the form. In addition, most hospitals use hazard stickers for this purpose. In this event, the specimen is usually left in the fixative for a longer period before being examined and cut up in a class I safety cabinet. Culture swabs may be taken from the fresh specimen by the surgeon before fixation and sent to the microbiology department.

Cut-up

All specimens are measured and their macroscopic features described. Solid organs are weighed. Biopsy material is put through for processing intact. The rest are cut up and representative pieces of the tissue processed. Specimens of organs are best left uncut (solid) or unopened (hollow) and simply immersed in the fixative solution, providing they arrive at the pathology department on the same day. The pathologist can then cut or open the specimen, clean it if necessary and thereafter, in the case of hollow organs, pin it out on a cork board. The material is then immersed in fresh fixative for representative blocks to be taken on the following day. This ensures good preservation of the interior of the specimen or mucosal lining, avoids distortion and permits better histological evaluation. One exasperated pathologist, upset by the persistent habit of surgeons in opening the specimens, was known to reiterate that 'surgeons should get out of the habit of opening their bowels in theatre'.

The removal of pieces of fresh tissue for any purpose, e.g. estimation of oestrogen receptors or research, is best undertaken by the pathologist, who can ensure that the correct pieces are given while retaining sufficient material to enable a firm histological diagnosis. Gross specimens, if considered to be of special interest, are photographed before blocks are taken for processing and histological analysis.

Tissue processing, cutting and staining

Processing and embedding

The aim of tissue processing is to embed the fixed tissue in a supporting medium that will permit the subsequent cutting of histological sections without deformation or damage of the tissue. In practice, paraffin wax is the most satisfactory routine embedding medium. The actual steps involved after fixation include:

- dehydration to remove the fixative and the tissue water by alcohol and acetone
- clearing of the tissue by substances that are miscible with the dehydrating agents and the embedding medium, i.e. xylene or toluene
- embedding in hot wax.

Nowadays, the dehydration and clearing of the fixed tissue are conducted by the automatic tissue processor.

Cutting and staining

Sections are cut on a microtome and are then mounted on glass slides before staining. The routine stain used is H&E, and this provides enough details of cellular and tissue architecture to enable the vast majority of routine histopathological diagnoses to be made. However, there are situations in which special stains, including histochemical staining (Figures 1.27 and 1.28), are needed to identify specific tissue components or cellular inclusions.

Reporting

Laboratories do vary, but usually a report should be available to the clinician within 2–3 days. If an urgent report is required, the pathologist can give the specimen priority and examine it before the rest, thus enabling issue of an early report. This practice, which is in widespread usage, must not be abused, as this would render impossible the organization of a priority service by the pathologist. The request for an 'urgent paraffin' report must be clearly indicated on the form and telephone or 'bleep' details of the contact doctor included.

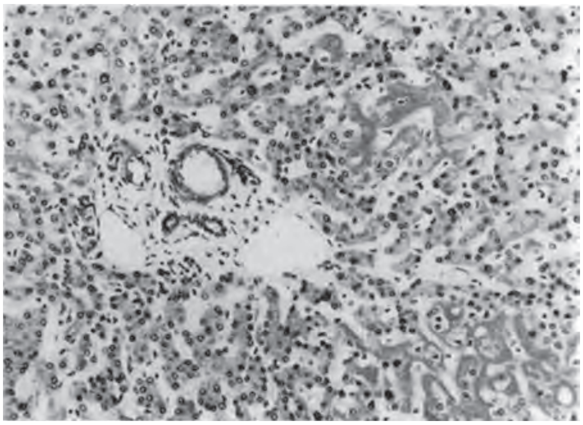


Figure 1.27 Liver stained by Congo red to demonstrate deposition of amyloid.

Although routine reports are issued within 24–48 hours, this is not always possible for a variety of reasons. Some tissues require longer to fix (infectious) or process (fatty tissue, breast) or require special treatment, e.g. decalcification (bone), before processing. Further material may need to be examined and deeper levels cut or special staining or electron microscopy carried out before a firm pathological diagnosis can be reached. Finally, difficulties may be encountered with pathological interpretation that may require consultation and, in some cases, outside specialist opinion. However, in these situations, it is usually possible for the pathologist to issue an interim report.

Frozen section

In this technique, the fresh tissue is frozen (-25°C) using solid CO_2 or liquid nitrogen and then sectioned in a special cabinet (cryostat) containing a microtome. The cut frozen sections are stained, usually with H&E, for immediate reporting to the surgeon. It is an exacting technique that requires the services of an experienced histopathologist, as the interpretation is more difficult than that experienced with fixed paraffin sections. Furthermore, there is some pressure on the pathologist to provide the information needed in a short time interval. In all cases, it is essential that the pathologist be briefed with the details of the clinical features and the nature of the biopsy taken. If all of these conditions are fulfilled, the accuracy of frozen section reporting is high, with low false-positive (0.2%) and false-negative (0.6%) results.

Frozen section is used for three purposes during surgery:

1. To establish the nature of suspect lesions: is it benign or malignant?
2. To ensure disease-free margins in resections for malignancy.
3. To establish whether the biopsy obtained in a difficult case contains sufficient material for a diagnosis to be established with subsequent paraffin histology.

Although the pathologist is able to give a definite diagnosis in the majority of cases, there are instances when this is not possible

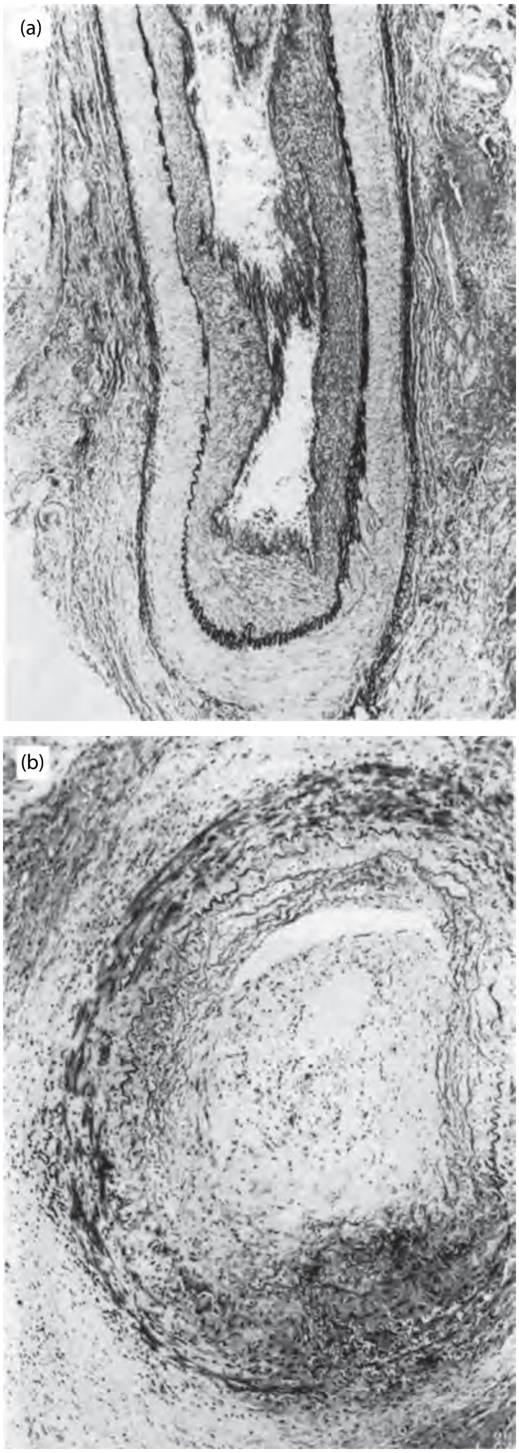


Figure 1.28 Elastica red/yellow stain: (a) normal artery; (b) disruption of elastic tissue in giant cell arteritis.

and the diagnosis has to be deferred until permanent sections are examined. In this situation, the surgeon, after ensuring that the pathologist has received adequate tissue, may opt to terminate the procedure and postpone definite surgical treatment until the result of the paraffin histology is available.

All frozen sections in patients undergoing elective surgery should be booked with the pathology department prior to the operation. There is no doubt that the frozen section service is abused by surgeons and a large number (50% in one reported audit) are unnecessary. *The most important guideline to the surgeon in deciding the need or otherwise for a frozen section is whether the findings of this process will influence the nature of the procedure carried out at that time.*

Handling of special specimens

Breast screening

Breast screening programmes are based on mammography performed in screening units or mobile vans. If an abnormality is detected on the initial screen, the woman is called for more detailed mammography and clinical examination. Some of these individuals are found to have palpable breast lumps. Others have impalpable lesions, which appear on the mammographic films as 'areas of disordered architecture', 'microcalcification' or 'localized densities', or a combination of these. These lesions require preoperative localization by special techniques (most commonly by the Mammolock needle inserted by the radiologist) immediately prior to excision biopsy performed under general anaesthesia (see Chapter 17).

The specimen with the localizing needle is radiographed in the pathology department using a Faxitron machine. These specimen films are compared with the preoperative mammographic films to determine whether the lesion has been removed. If the abnormality is not identified, the surgeon is informed and more tissue is removed, and the process repeated until radiographic confirmation of excision of the lesion is obtained. In practice, most lesions are identified on the first biopsy (Figure 1.29).

The circumference of the specimen is then marked with dye. Following inspection and palpation, the specimen is cut in serial coronal slices to a predetermined thickness (5.0 mm) on a special board. The tissue slices are then placed in sequence on a radiographic film, which is then exposed to determine the exact location and extent of the lesion (Figure 1.30). Blocks of tissue are taken from the slices containing the radiographic abnormalities and are labelled to match the slices from which they are taken. Particular attention is paid to the resection margins. The blocks are then fixed in formalin and processed for routine paraffin histology.

Approximately 50% of the subclinical lesions identified by screening are benign (sclerosing adenosis, duct ectasia, etc.) and the rest malignant. Non-invasive cancer (ductal carcinoma *in situ*, lobular carcinoma *in situ*) is more commonly found than invasive cancer in these screening biopsies. Compared with the naturally presenting disease, a greater incidence of tubular and cribriform invasive cancer is found by breast screening.

Lymph nodes

Ideally, these should be received fresh and are handled by the pathologist in a safety cabinet. The node is bisected and dabs (imprint cytological smears) are made on glass slides. These are stained with Giemsa and Lieshman stains. Thereafter, the procedure is as follows:

- Representative cross-sections of the nodes are fixed in formalin for paraffin embedding and routine H & E staining.
- Another piece is fixed in formalin but is embedded in resin to enable cutting of semithin sections (1–2 μm), which give better morphological details of lymphoid cells (Figure 1.31).

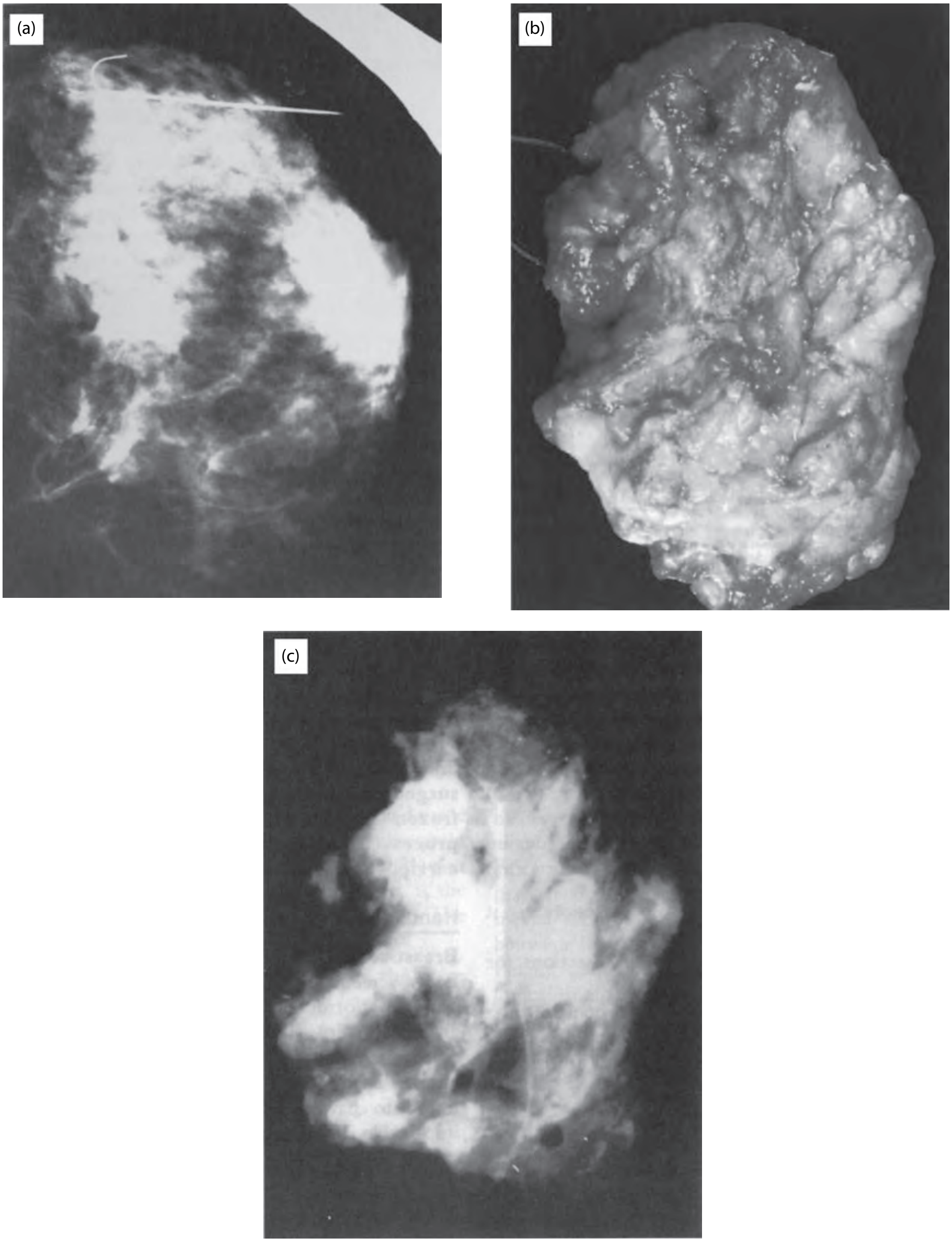


Figure 1.29 (a) Localizing mammogram with a Mammalock needle; (b) excisional biopsy of the lesion; (c) radiograph of the same specimen.

- Small cubes (2.0 mm^3) are fixed in glutaraldehyde for electron microscopy.
- A piece is snap frozen (−70°C) for enzyme and immunohistochemistry. The latter is important for the identification of T- and β-cells using specific monoclonal antibodies.
- A small piece is kept in the refrigerator at 4°C for microbiological examination should this be desirable: the infective nature of the lesion may not have been appreciated at the time of the biopsy. In this situation, culture of the retained fresh lymph node may furnish valuable confirmatory evidence for the establishment of a definite diagnosis.

Small intestinal biopsy

The specimen is removed from the capsule and dabs made onto glass slides, which are then stained with Giemsa or toluidine blue in the search for *Giardia lamblia*. The specimen is then placed in a Petri dish containing isotonic saline and examined under the dissecting microscope to assess the villous architecture. The orientation of these small specimens before embedding is important for accurate histological interpretation. If enzyme histochemistry is desired a portion is kept deep frozen.

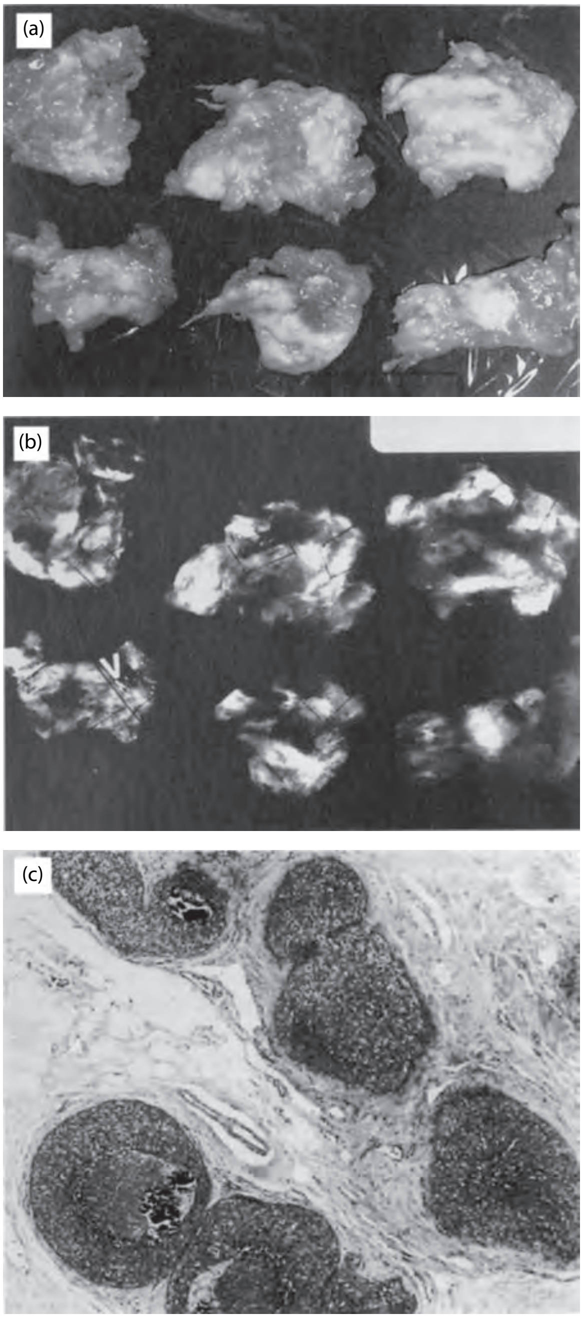


Figure 1.30 (a) Slices of the specimen; (b) radiograph of the specimen with microcalcification (arrow) in one of the slices; (c) histology of the lesion from this slice (ductal carcinoma *in situ*).

Skin biopsies

Most skin biopsies are fixed in formalin as the purpose of the biopsy is usually to exclude malignancy or confirm the exact diagnosis of benign lesions. However, when certain skin disorders are suspected, e.g. chronic autoimmune vesiculobullous disorders, all forms of lupus erythematosus, leucocytoclastic vasculitis, etc., the skin biopsies are sent fresh. These are processed in a special manner. The skin ellipse is divided longitudinally and one half is frozen for direct immunofluorescent examination to detect IgA, IgG, IgM and C3. The other half is fixed in formalin for routine paraffin blocks and H&E staining.

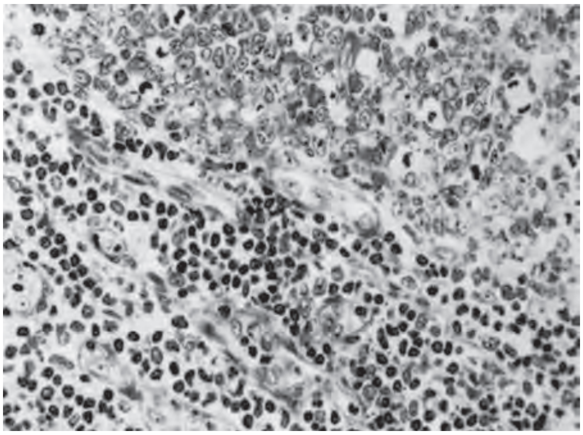


Figure 1.31 Semithin section showing greater cellular detail of the lymph node.

Muscle biopsies

The specimen should be received fresh so that special procedures can be carried out and the correct orientation obtained. The specimen is divided into three portions. Small blocks are fixed in glutaraldehyde for electron microscopy. A representative cross-section is snap frozen for diagnostic enzyme histochemical studies. These include staining with Gomori trichrome, periodic acid–Schiff with and without diastase, ATPase at various pH values, NADH diaphorase, non-specific esterase and SR19. A myophosphorylase is performed if glycogen deficiency is suspected. The remainder is left to relax on a piece of cardboard for 10 minutes before being cut in longitudinal and transverse blocks for fixation in formalin, paraffin embedding and routine H&E staining.

Enzyme histochemistry

As enzyme activity is partially or totally lost when tissue is fixed and processed in paraffin blocks, enzyme histochemistry is usually performed on fresh-frozen material. These techniques are not employed routinely but are indicated in certain specific circumstances, the most important of which are:

- myopathies: in which a specimen of skeletal muscle is stained for the following enzymes involved in normal muscle metabolism and contraction: ATPase (which differentiates type I from type II fibres; Figure 1.32), phosphorylase, cytochrome oxidase, phosphofructokinase and aldolase
- malabsorption: for the diagnosis of alactasia, a jejunal biopsy is examined for lactase and sucrase deficiency, whereas in suspected gluten enteropathy the jejunal biopsy may be stained for acid and alkaline phosphatase
- diagnosis of Hirschsprung disease: rectal biopsy is stained by the acetylcholinesterase method for the identification of nerves and ganglia
- detection of acid phosphatase in metastatic prostatic cancer
- detection of alkaline phosphatase in vascular endothelial tumours.

Immunohistochemistry

This is a technique by which specific antibodies are utilized to detect cellular or tissue constituents. Direct techniques involve the application of the specific antibody to the tissue.

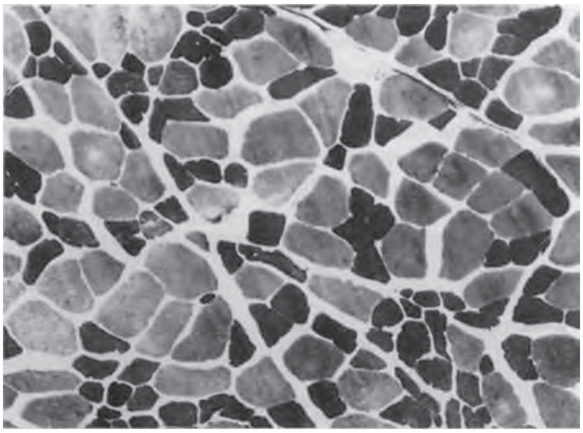


Figure 1.32 Muscle biopsy stained with ATPase showing type I and type II fibres.

As the antibody is labelled (by conjugation with a marker), the presence and site of the antigen–antibody reaction can be determined (Figure 1.33). These direct methods are, however, insensitive. They have been largely superseded by indirect (sandwich) immunohistochemical techniques that are much cleaner and more sensitive. The indirect method employs two antibodies. The primary antibody (for the detection of the desired antigen) is unlabelled and is applied to the tissue. The second tracer antibody raised in a different animal species against the primary antibody is labelled. When applied to the section it will bind to the primary antibody and the resulting complex (tracer antibody + primary antibody + antigen) is detected by the marker on the tracer antibody. The most sensitive indirect method is the avidin–biotin complex technique, in which use is made of the fact that avidin (derived from egg white) binds with high affinity four molecules of biotin (a water-soluble vitamin). The labelled biotin is conjugated to the tracer antibody (biotinylated). The first stage consists of the application of the mouse primary monoclonal antibody to the tissue for binding to the target antigen. In the second stage, the labelled biotinylated anti-mouse tracer antibody is added. This binds to the primary antibody. Finally, the addition of avidin-labelled biotin complex results in the formation of the avidin–biotin complex binding the entire conglomerate by strong molecular bridging (Figure 1.33b). This technique has been modified with the substitution of avidin with streptavidin (protein component of *Streptomyces avidinii*), which is less prone to non-specific attachment to tissue lectins.

The most common markers or labels used as conjugates with the tracer antibodies are enzymes (horseradish peroxidase, calf intestinal phosphatase, glucose oxidase), since these can be readily identified by specific colour reactions with chromogenic substrates. Other markers include colloidal metals (gold, silver, silver-enhanced gold, ferritin), radioisotopes and fluorescent labels. Radiolabels are used for quantitative immunocytochemistry and autoradiography.

The most important significant advance to contribute to the widespread routine use of immunohistochemistry was the development of the hybridoma technique for the production of monoclonal antibodies. It consists of the fusion of a normal plasma cell or transformed B-lymphocyte with a neoplastic myeloma cell. The resulting hybridoma cells are immortal and, during in vitro cell culture, can produce large quantities of a variety of antibodies when exposed to antigen. By careful subculture the clone producing the desired antibody that is specific to the antigen in question and which does not react with any other molecules is isolated. The resulting pure antibody is referred to as monoclonal (derived from one clone of cells) to distinguish it from polyclonal antibody, i.e. by several clones of plasma cells, as happens when a host animal is immunized with a purified specific antigen.

In general, the fixation of tissue does not destroy the antigens of interest to diagnostic histopathology and thus immunohistochemistry can be used on fixed tissues. There are, however, some exceptions. The important ones include lymphoid cell-surface antigens, cell membrane or nuclear receptors, hormones and neuropeptides. These are best detected on freeze-dried specimens or frozen sections.

In practice, immunohistochemistry is important in the following situations:

- Typing of lymphoid tumours with the use of antibodies that recognize specific surface antigens on leucocytes. These are referred to by the CD system and are used to classify lymphomas into β-cell origin (CD45R), T-cell origin (CD3), Hodgkin's (CD30, CD15) and histiocytic lymphomas (CD68).
- Identification of specific tumours: (1) germ cell tumours by placental alkaline phosphatase (seminomas, dysgerminoma), α-fetoprotein (non-seminomatous germ cell tumours, embryonal carcinoma), human chorionic gonadotrophin (HCG) (malignant teratoma trophoblastic), human placental lactogen (HCG-positive germ cell tumours); (2) carcinoid tumours by Singh (argentaffin) and Grimelius (argyrophil) (Figure 1.34); (3) prostatic tumours by prostate specific antigen (Figure 1.35), prostatic acid phosphatase; (4) melanoma by S100, HMB45 and vimentin; (5) thyroid tumours by thyroglobulin for thyroid epithelial lesions and calcitonin for medullary carcinoma; (6) vascular tumours by endothelial markers – factor VIII (CD34); and (7) early lymph node metastatic deposits (Figure 1.36).
- Soft-tissue tumours: in these tumours, immunohistochemistry is used to identify specific intermediate filaments such as vimentin, actin and desmin. The last is found in skeletal and smooth muscle cells (normal and neoplastic). Actin identifies myoepithelial cell origin.
- Anaplastic tumours: some tumours are so anaplastic that ordinary histological examination cannot even identify the tissue of origin of the tumour: epithelial, soft tissue or lymphoma. This problem can be resolved by immunohistochemical staining with a panel of monoclonal antibodies. The first step in the solution to this problem is the differentiation of the tissue of origin: epithelial, lymphoid or mesenchymal. Epithelial origin is identified by a panel of antibodies to cytokeratins, epithelial membrane antigen and carcinoembryonic antigen, lymphomatous origin by leucocyte common antigen, mesenchymal origin by vimentin and desmin (Table 1.3).

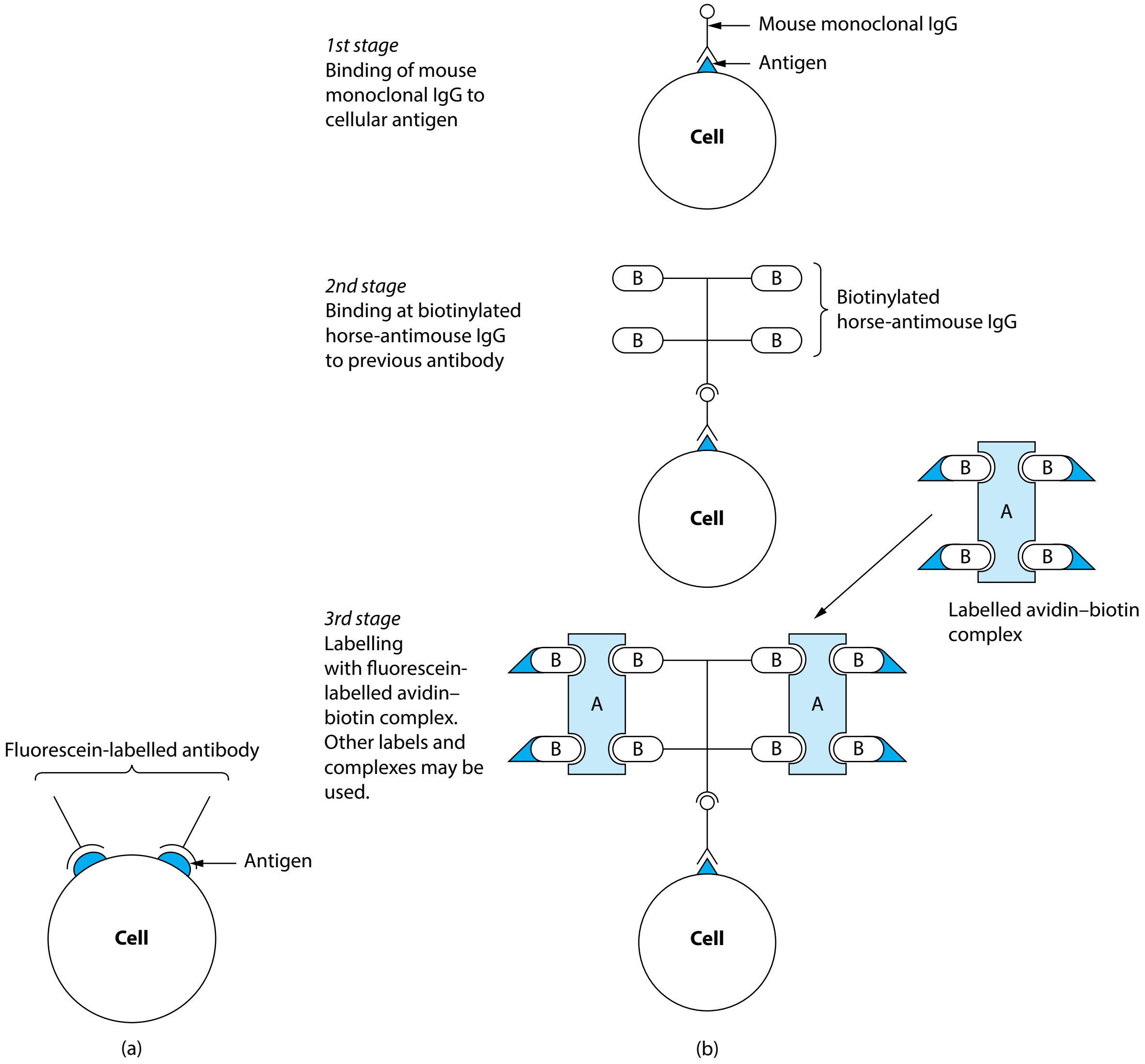

1st stage
Binding of mouse monoclonal IgG to cellular antigen
Mouse monoclonal IgG
Antigen
Cell
2nd stage
Binding at biotinylated horse-antimouse IgG to previous antibody
B
Biotinylated horse-antimouse IgG
Cell
A
Labelled avidin–biotin complex
3rd stage
Labelling with fluorescein-labelled avidin–biotin complex. Other labels and complexes may be used.
Fluorescein-labelled antibody
Antigen
Cell
(a)
(b)

Figure 1.33 Principles of immunohistochemical staining: (a) direct technique; (b) indirect technique.

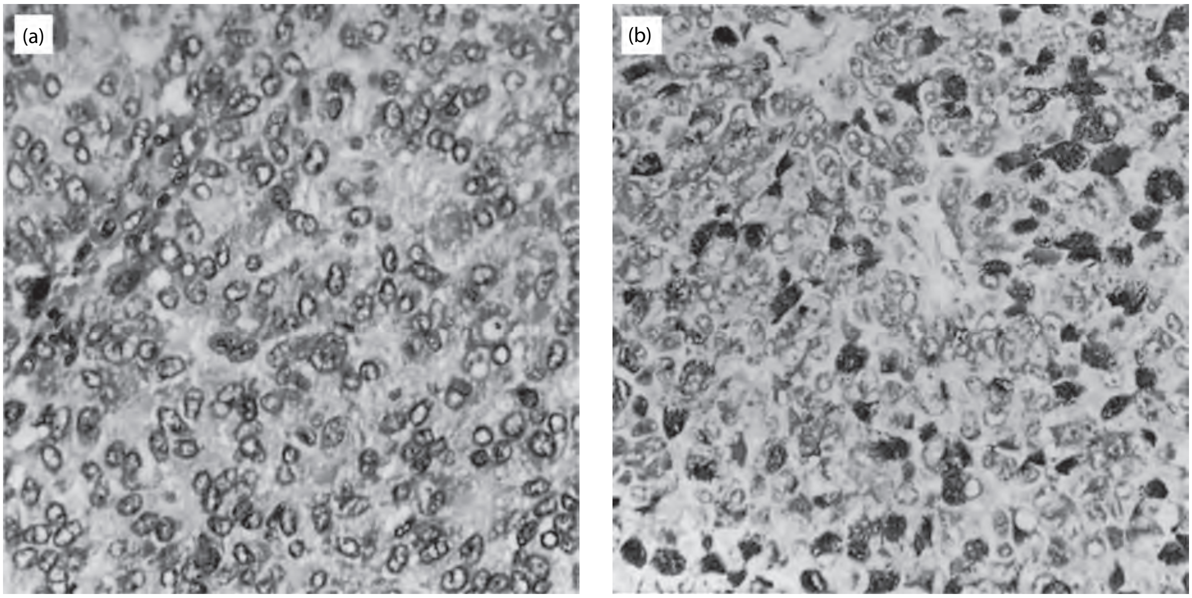


Figure 1.34 Carcinoid tumour: (a) haematoxylin and eosin staining; (b) Grimelius staining showing positive argyrophilic cells.

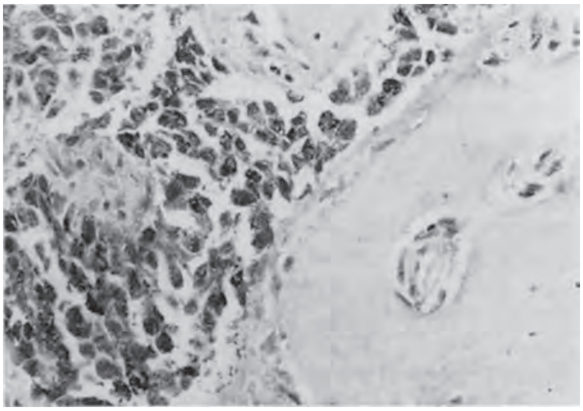


Figure 1.35 Bone biopsy. Positive staining of the tumour cells for prostate-specific antigen.

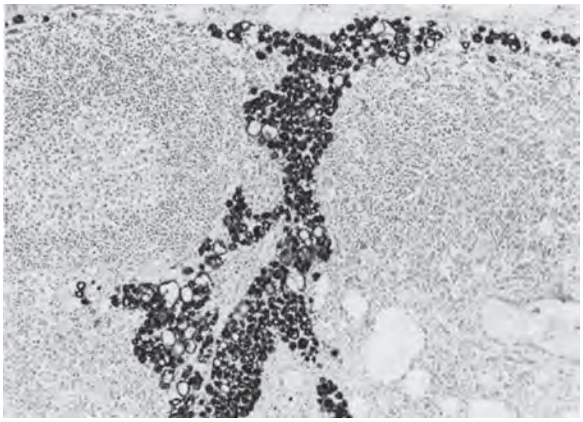


Figure 1.36 Lymph node stained to reveal cytokeratin-positive cells (metastatic breast carcinoma).

Table 1.3 Immunohistochemical characterization of anaplastic tumours

Antibody against	Nature of tumour
Cytokeratins	Epithelial
Epithelial membrane antigen	Epithelial
Leucocyte common antigen	Lymphoid
Vimentin	Mesenchymal, lymphoid
Desmin	Muscle
Actin	Myoepithelial
Neuron-specific enolase	Neuroendocrine, neuronal
S100	Melanocyte, nerve sheath
Factor VIII	Endothelial
CD34 (QB end)	Endothelial

Immunofluorescent techniques

Fluorochromes (fluorescein isothiocyanate isomer I, tetramethyl rhodamine isothiocyanate isomer R) are used as labels for the tracer antibody. A fluorochrome absorbs radiation in the ultraviolet or visible light range. As it is bound irreversibly to the antibody, the site of the antigen–antibody complex is visible by fluorescence even at the subcellular level, e.g. mitochondria, microsomes and muscle fibres. This technique is often used in flow cytometry (see below).

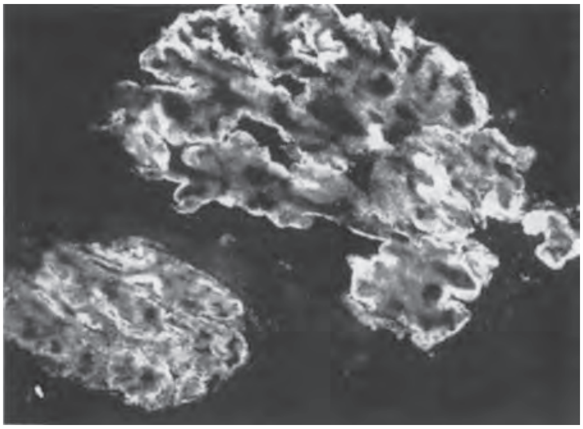


Figure 1.37 Renal biopsy: fluorescent staining of a frozen section specimen showing positive immunoglobulin G in membranoproliferative glomerulonephritis type I.

The main routine indications for immunofluorescent staining are skin and renal biopsies. In the former, they serve to outline the patterns of immunoglobulin distribution in systemic lupus erythematosus, bullous disorders and the vasculitides. In renal biopsies immunofluorescent techniques are used to localize immunoglobulins, complement and fibrin within the glomerular basement membrane, mesangium and vessel walls. Various renal disorders exhibit different specific distribution of these substances (Figure 1.37).

Electron microscopy

Electron microscopy uses an electron beam produced by an electron gun. There are two types of instrument: transmission and scanning electron microscopes, although modern versions of the former have a scanning capability. Special fixatives (e.g. osmium tetroxide, glutaraldehyde), embedding media (e.g. epoxy resin) and stains (e.g. lead citrate, uranyl acetate) are used for electron microscopy.

Transmission electron microscopy

The principle is similar to that of light microscopy. However, instead of visible light a high-velocity homogeneous beam of electrons is focused on the specimen that is thin enough to transmit at least 50% of the incident electrons. The transmitted beam then enters the imaging system composed of a series of lenses, which together produce the final, highly magnified image of the specimen (Figure 1.38). Modern instruments can achieve a resolution of 0.14 nm.

Scanning electron microscopy

When an electron beam hits a specimen, some electrons are transmitted or back scattered, whereas others interact with the atoms of the specimen to produce secondary electrons (low energy), X-rays and visible light (cathode luminescence). The secondary electrons are used to form three-dimensional images of the surface architecture of the specimen when the instrument is in the emissive mode (Figure 1.39). In the X-ray mode, scanning electron microscopy (SEM) can be used to provide a chemical analysis of the specimen.

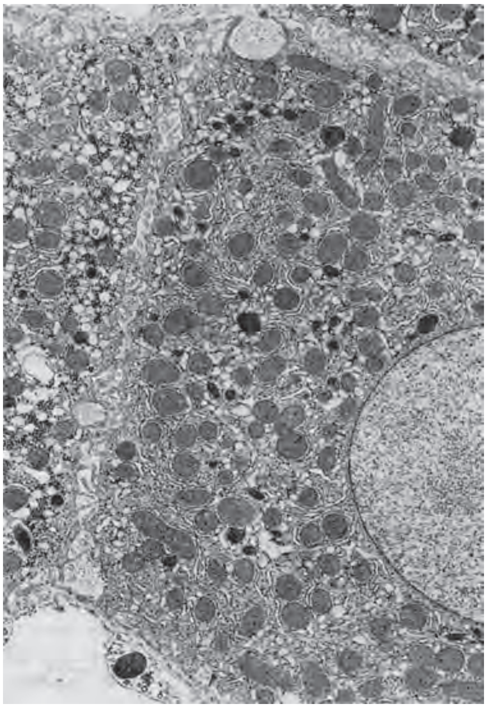


Figure 1.38 Transmission electron microscopy: liver hepatocyte.

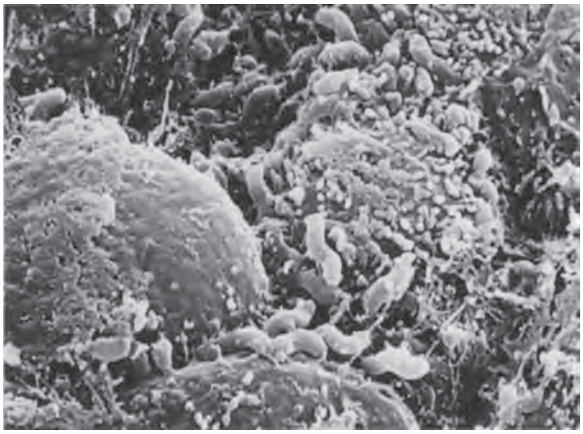


Figure 1.39 Scanning electron microscopy: *Helicobacter pylori* in gastric mucosa.

Diagnostic use of electron microscopy

In routine practice, electron microscopy is used for the following procedures:

- Interpretation of renal biopsy material: for the identification of immune complex deposits in basement membranes and mesangium (Figure 1.40).
- Tumour diagnosis: electron microscopy is used to differentiate an anaplastic carcinoma from a lymphoma, in the identification of neurosecretory granules in APUD (amine precursor uptake decarboxylase) tumours (e.g. carcinoids, medullary carcinoma of thyroid), and identification of melanosomes in suspected melanomas and in the differentiation of spindle cell tumours and lymphomas. Even with histochemistry, spindle cell tumours can be difficult to interpret. In these tumours, electron microscopy can help to differentiate epithelial (spindle cell carcinoma) from smooth muscle or fibroblastic tumours. Although electron microscopy can be used to identify the cell origin of lymphomas, this is better achieved nowadays with immunohistochemistry.
- Diagnosis of inborn errors of metabolism.
- Diagnosis of viral and other infections.
- Skin disorders: electron microscopy is useful in the diagnosis of viral disorders, skin infiltration by T-cell lymphomas, classification of epidermolysis bullosa and histiocytosis X, etc.
- Identification of amyloid.

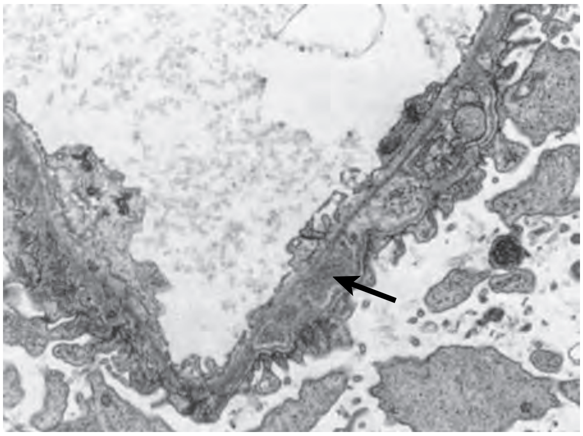


Figure 1.40 Renal biopsy: deposits in the basement membrane (arrow) in membranous glomerulonephritis.

Quantitative techniques

Histometry and stereology

The techniques of histometry provide quantitative measurements of histological preparations such as volume proportions, surface area, particle/tissue component numbers and particle dimensions. The techniques involve the use of purpose-designed glass covers, which incorporate parallel lines (straight or sinusoidal) or spatially related dots (e.g. outlining equilateral triangles) on the mounted specimens. The number of particles, cells, etc., under study that are crossed by the lines or superimposed by the dots is counted (point counting). Histometric techniques have a number of applications:

- measurement of certain histological changes: cellularity, staining intensity and grading of tumours
- examination of three-dimensional features from the two-dimensional planes using geometrical procedures (stereology), e.g. number of cells per volume, surface area per volume
- valid comparisons of histological studies from different centres because the results can be standardized.

Flow cytometry

This technology is in established usage for the following:

- rapid measurement of the biological properties of cells or subcellular organelles
- physical separation of desired subpopulations of cells based on these measured properties
- objective measurements of cellular features, including tumour cell DNA analysis (e.g. ploidy)
- cell cycle analysis.

Flow cytometry is applicable to blood, bone marrow, malignant effusions, lymphoid tissues and cell suspensions from solid tumours, although the last are more difficult to prepare and require extensive mechanical or enzymic disaggregation. The flow cytometric work published during the 1990s has led to a better understanding of the behaviour and classification of many tumours. The main application of flow cytometry has been in the study of solid tumours and cellular immune response to tumours and organ allografts. Flow cytometric examination of tumours can support the diagnosis when morphological examination is equivocal and has provided useful data in the study of borderline lesions. In addition, the ploidy status of a tumour can provide useful prognostic information, which is independent of stage and grade. It has been used to monitor responses to chemotherapy and to evaluate tumour recurrence after successful treatment.

The principle of flow cytometry entails the passage of a single column of cells in suspension through a transparent chamber. The cell column is irradiated by a beam of laser light that is scattered at various angles depending on the characteristics of the cells and their inclusions. The scattered light is detected and converted to electronic signals, which are digitized and stored by an online computer that is able to generate the required information in histogram form, e.g. cell size, viability and cell granules. Specific stains are used according to the intended investigation. For tumour ploidy and proliferative activity a DNA fluorochrome (propidium iodide) is used to stain nuclear suspensions obtained from the tumour. The flow cytometric profile of normal cells consists of a predominant peak in the diploid (2N) region and a much smaller peak in the tetraploid (4N) region. The former corresponds to cells in the G0–G1 phase (resting) and the latter to the G2–M phase (post-DNA synthesis/mitosis) of the cell cycle. Tumour cells exhibit additional (abnormal) aneuploid peaks with nuclei containing abnormal amounts of DNA relative to the diploid state. The degree of aneuploidy is defined by the DNA index, which is the ratio of the DNA content of the G0–G1 aneuploid peak to the DNA content of the G0–G1 diploid peak. For most tumours, this is greater than 1.0 (hyperdiploid) and, much less commonly, lower than 1.0 (hypodiploid).

Flow cytometry is also used to provide data on the cell cycle kinetics of tumours. With the aid of specific computer programs and specialized techniques (bromodeoxyuridine uptake, Ki-67 monoclonal antibody), the relative proportions of cells in the G0–G1, S1 and G2–M phases can be calculated. In general, DNA aneuploidy and elevated proportions of cells in the S phase (synthesizing DNA) indicate aggressive tumours with a poor prognosis.

Autoradiography

In autoradiography a radioactive label (β-emitting isotope) such as tritium (^{3}H) or ^{125}I is introduced into the tissue either by injecting an animal or following *in vitro* incubation of tissue or cells. The labelled tissue is then opposed to a special water-soaked film (Kodak AR 10 stripping film) or dipped in special photographic emulsions (Eastman-Kodak, Ilford, UK). The autoradiographic process results from the activation of the emulsion by the β-radiation from the radioisotope as it decays. The resulting image reproduces in great detail the distribution of the radioactive label within the section.

Autoradiography is rarely used for routine purposes but is an extremely useful tool in biological research, including studies involving *in situ* hybridization for the detection of DNA and RNA sequences with radiolabelled probes.

Hybridization techniques

In the production of cellular polypeptides, the genetic instructions contained in the double-stranded DNA genome are copied on to the single-stranded messenger RNA (mRNA) by a process known as transcription, with the DNA acting as the template. The message encoded in the mRNA is then used in the process of translation to form the specific polypeptides. The detection of abnormal DNA or mRNA can be achieved in the laboratory in two ways. The first consists of immobilization by heating of preparations of the DNA or RNA on inert supporting membranes (nitrocellulose or nylon), followed by the addition of specific labelled (radioactive or chromogenic) nucleic acid probes (produced by recombinant biotechnology),which then hybridize with the specific sequences. In the case of DNA, denaturation by alkali is essential to convert the molecule into a single strand prior to the application of the specific probe. Examples of filter hybridization (or blotting techniques) are dot–blot hybridization for measuring specific amounts of DNA or RNA extracted from tissues, Southern blotting for the detection of DNA fragments produced by digestion using special enzymes called restriction endonucleases, northern blotting for the determination of the molecular size of specific mRNAs and western blotting for the detection and sizing of specific polypeptides.

The second technique involves the demonstration of the abnormal DNA and RNA in the tissues (formalin fixed or fresh) and is known as *in situ* hybridization. The process is essentially similar to that involved in the blotting techniques. The cellular DNA is first denatured by heat or alkali, specific labelled nucleic acid probes are applied in the presence of formamide and 1.2 M NaCl, and the specimen is then incubated for several hours. Non-specific staining is minimized by using the smallest amount of labelled probe possible and an excess of unlabelled non-competing DNA or RNA to block non-specific binding sites.

Although *in situ* hybridization is likely to play an increasing role in histopathology in the future, at present its use is restricted to research institutions and major centres. Currently, it is employed in the detection of viral colonization (e.g. cytomegalovirus, human papillomavirus, Epstein–Barr virus), examination of genomic material in hereditary disorders and clonal derivation of lymphomas.

GUIDE TO FURTHER READING

Chin G, Diegelman R, Schultz G. Cellular and molecular regulation of wound healing. In: Falabella A, Kirschner R (eds) *Wound Healing*. Boca Raton, FL: Taylor, Francis Group, 2005:17–37.

Edge SB, Byrd DR, Compton CC, *et al.* (eds). *AJCC Cancer Staging Manual*, 7th edn. New York, NY: Springer, 2010.

Quirke P, Steele R, Monson J, *et al.*; MRC CR07/NCIC-CTG CO16 Trial Investigators; NCRI Colorectal Cancer Study Group. Effect of the plane of surgery achieved on local recurrence in patients with operable rectal cancer: a prospective study using data from the MRC CR07 and NCIC-CTG CO16 randomised clinical trial. *Lancet* 2009;**373**:821–828.

Roitt IM. *Essential Immunology*, 11th edn. Oxford, UK: Wiley-Blackwell, 2006.

Tilney HS, Rasheed S, Northover JM, Tekkis PP. The influence of circumferential resection margins on long-term outcomes following rectal cancer surgery. *Dis Colon Rectum* 2009;**52**:1723–1729.

Westra HS, Hruban RH, Phelps TH, Isacson C. *Surgical Pathology Dissection: An Illustrated Guide*, 2nd edn. New York, NY: Springer, 2002.

CHAPTER 2

Surgical craft approaches and technologies

SIR ALFRED CUSCHIERI

Introduction

Surgeons require (1) knowledge of the disease processes that they manage; (2) cognitive skills that underline clinical judgement and decision-making; and (3) the necessary level of psychomotor skills to ensure technical operative competence – the ability to perform operations consistently well hence ensuring safety of the patient; therefore, surgery is one of the *craft specialties.* All of these components of surgical practice are important, but the level of technical competence distinguishes the master surgeon from the average. In psychomotor research, ability is defined as 'the adaptive capacity, trait, attribute or aptitude that an individual brings to a given task', whereas skill is the result of a specific combination of abilities reinforced by practice and training. Ability is thus innate and determines the ultimate level of skill that an individual will attain with practice and experience. Performance relates to the overall efficiency with which a complex activity such as a surgical operation is executed. Competence is reached when an individual can perform a task safely, to an acceptable standard and within an acceptable time frame, i.e. the individual can perform the operation safely and with good patient outcome, but is still acquiring experience in executing the procedure and has to use what the ergonomic psychologist calls 'control active mode' throughout. Proficiency indicates that competence is consistent and has reached the 'expert level' as a result of experience and reinforcement by a sustained caseload. This is the level that every surgical trainee must achieve as an independent consultant. In achieving proficiency, the surgical trainee progresses though a proficiency-gain curve that is as specific to an operation as it is to the individual trainee. Proficiency gain requires supervised training by a tutor assisting the trainee and reinforcement, i.e. perform the same operation (caseload effect) until the proficiency plateau (Figure 2.1) is reached, when surgeons attain the semiautonomic operating mode, i.e. they can do the operation without having to think of the various steps of the procedure. Unfortunately, the *proficiency-gain curve* is often erroneously called the *learning curve*, but the two are not equivalent. Learning refers to acquisition of knowledge of the indications for the operation, the component tasks and steps of the procedure, the instruments and technologies

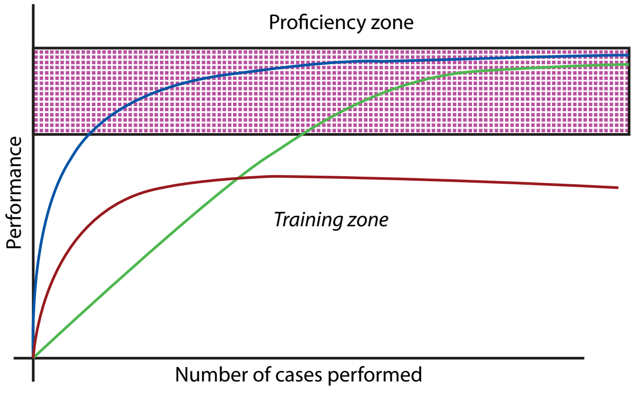


Figure 2.1 Proficiency-gain curve. The proficiency zone refers to the proficiency plateau that the surgical profession and society expects from fully trained surgeons certified and credentialed or licensed to operate on patients. For most trainees (the green line), the proficiency plateau is reached after doing the operation under supervision *x* number of times (reinforcement/case load effect), with *x* varying with the individual trainee (some need more than average, others less) and the operation (minor, medium, advanced, specialist). The blue line refers to individuals who are innately very gifted with the necessary psychomotor skills, such that they reach the proficiency plateau much quicker, requiring less cases. The red line denotes individuals who, despite having the required cognitive skills, lack the necessary level of innate psychomotor ability to be able to reach the proficiency plateau. It is the job of the programme director to identify these trainees at an early stage and advise on redirection of their medical careers.

required to do the operation, etc., but gives no indication that the individual can actually do the operation.

Until recently, operative training was based exclusively on the apprenticeship system. This entails graded exposure to operative work depending on progress as assessed by the consultant. The trainee graduated from second to first assistant, to performing minor operations under supervision by more senior surgeons, followed by exposure to major surgery initially as first assistant, then as principal surgeon under supervision until competence to perform certain operations unsupervised is reached. Competence is, of course, specialty and procedure related. Increasingly, the apprenticeship system is being supplemented by training of component operative generic skills, e.g. suturing and anastomosis, in purpose-designed skills laboratories or units. This has assumed greater importance since the period of surgical training was reduced.

Categories of operation

Operations can be classified by various criteria:

- total operative insult to the patient
- risk of infection by the procedure
- urgency of treatment
- lack of progress of acute disease
- remedial
- ablative or functional
- prophylactic
- validated.

Total operative insult

An operation, however expeditiously performed, inflicts injury to the patient and can really be defined as a controlled traumatic insult on a patient with a therapeutic intent. The total operative trauma has two components: (1) the trauma of the exposure needed to reach the operative site and (2) the trauma of the procedure itself (procedural trauma). There is no correlation between the access trauma and the procedure trauma. Indeed, with some operations, the access trauma accounts for the major part of the total operative insult, e.g. vagotomy or cardiomyotomy. The ratio of access trauma to procedure trauma of specific operations determines the benefit obtained by using the laparoscopic approach instead of open access by laparotomy. Benefit is likely to be significant if this ratio is high.

The extent of the total operative insult is used to classify operations into minor, medium and major, and it determines the rate of recovery of the patient from the intervention. The surgical trauma is responsible for the catabolic response and, together with general anaesthesia, for the non-specific immunosuppression. Although there is a relation between magnitude of the operation and the perioperative risks, it is important to stress that other factors are involved, chief among which is the preoperative state of the patient and overall quality of perioperative care.

Major operations require intensive or high-dependency care for this and other reasons, including cardiovascular and respiratory care and support. Patients undergoing medium operations require intensive or high-dependency postoperative care only if their preoperative condition is compromised by comorbid disease. Recovery from moderate operations includes a period of high-dependency care, followed by a stay in a surgical ward, discharge and convalescence before return to full activity or work. Convalescence is referred to in North America as the period of short-term disability (PSD). There is a wide variation in the PSD following major operations, as many factors operate in the individual case: personality type, social class, motivation, general health and performance status before the operation, type of healthcare provision (private or national), employment, etc.

Minor operations pose no immediate risk but nonetheless can be attended by complications. As the overall risk is minor, most of these patients are treated in day-case or ambulatory care surgical units.

Risk of infection

The majority of infections complicating surgical procedures are endogenous from the patient's own resident microflora. With good modern theatre practice, the risk of airborne infection is low and can be ignored except in prosthetic implant surgery and in the immune-compromised patient. The use of microenvironment modules and biological isolators is an established practice in joint replacement surgery and in patients at risk from opportunistic infections.

It is customary to classify surgical procedures into three categories depending on their infectivity risk:

1. *Clean operations*: do not involve bowel and are conducted in the absence of sepsis.
2. *Potentially infected (clean contaminated) operations*: include all elective operations on hollow viscera, which usually (low small bowel, colon), or on occasions, harbour pathogenic organisms, e.g. the biliary tract and upper gastrointestinal (GI) tract.
3. *Infected*: operations carried out for or in the presence of sepsis.

Ideally, clean operations should carry a negligible infection rate and any infection occurring in this category of operation indicates failure of the system and could jeopardize the long-term outcome, e.g. recurrence of hernia repair. The postoperative infection rate in the clean contaminated group is due largely to endogenous infection and can be substantially minimized by good aseptic technique, appropriate antibiotic prophylaxis and careful surgery designed to avoid spillage of the intraluminal contents with contamination of the peritoneal cavity and wound. The surgical principles governing the management of infected cases include treatment of the underlying pathology, efficient peritoneal toilet that entails complete evacuation of pus and culture of infected material. In addition, there is clear evidence for the value of two procedures in the management of severe intra-abdominal sepsis:

- peritoneal lavage with isotonic saline solution to which antibiotics may be added (tetracycline, cephalosporins)
- delayed closure of the abdominal wall: this usually refers to delayed closure of the skin and subcutaneous tissue layers on the premise that wound infection is likely; the musculoaponeurotic layer is closed with 1/0 monofilament absorbable material such as polydioxanone (PDS), and the wound is then packed with gauze soaked in proflavine

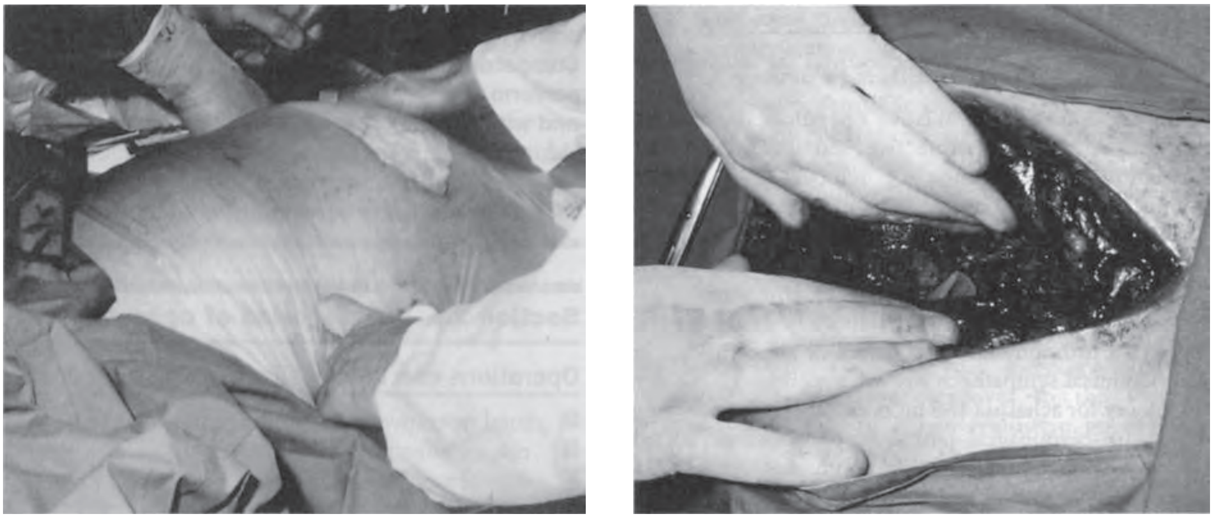


Figure 2.2 (a) Non-closure of the abdominal wall in severe sepsis (anterior coeliotomy, marsupialization). The patient had pancreatic necrosis with peripancreatic abscess formation. After resection of the pancreatic sequestrum, evacuation of the abscess and saline lavage, the space between the transverse colon and the stomach extending between the pancreatic bed and the parieties is packed with proflavine emulsion gauze. Evisceration is prevented by the application of an Opsite dressing. (b) Renewal of the pack and toilet of the cavity is performed every 24–48 hours under intravenous sedation in the operating theatre until sepsis is eliminated and healing is well under way.

The skin is closed 4–7 days later once granulation tissue has formed. If gross sepsis is encountered and recurrent intra-abdominal sepsis or abscess formation is considered likely, the entire laparotomy wound is left unsutured and the infected region of the peritoneal cavity and the wound are packed. Evisceration is prevented either by application of an Opsite dressing (Figure 2.2a) or by a prosthetic mesh (Prolene or polyester) fitted with a zip (or the two halves are sutured). The edge of each half is sutured to the edges of the skin wound. Irrespective of the method used, the peritoneal cavity is opened under sedation and the packing renewed at intervals of 1–2 days (Figure 2.2b), when the wound and the affected region are inspected and any necessary toilet and debridement carried out. In these severely septic patients, the antibiotic regimen must be based on the results of the bacteriological culture and sensitivity tests, and advice from the hospital microbiologist sought. Usually, the infection is a mixed one with both Gram-negative aerobes and anaerobes. Extra precautions are instituted in the operating theatre during and immediately after completion of infected cases to prevent the spread of infection. The theatre is thoroughly cleaned and the floor mopped with antiseptic solution before its further use.

Clean operations should precede potentially infected procedures in elective surgical lists. Infected cases are performed last, or preferably in a separate theatre reserved for these cases.

Urgency of treatment

Patients may require surgical treatment for the following reasons:

- They have chronic symptoms attributable to a diagnosed surgical disorder but have no acute manifestations or immediate life-threatening complications. In this instance, the operation is conducted on an elective basis. This category (elective operations) carries the lowest risk as there is time for detailed assessment and the patient is fully prepared physically, mentally and prophylactically for the planned operation. The underlying condition may, of course, be serious, e.g. cancer. In this situation, careful treatment planning on a multidisciplinary basis is essential in the individual case. In some instances chemotherapy or radiation therapy or both (chemoirradiation) is undertaken before the operation itself.
- Patients develop a life-threatening condition from trauma or because of an acute complication of an underlying benign or malignant disorder. These patients require emergency operation after adequate resuscitation (if needed). For all non-cardiac operations, the single most important risk factor is an emergency intervention. Thus, all emergency operations carry a definite risk of death but this varies with the nature, extent and severity of the underlying pathology, the cardiorespiratory reserve and age of the patient. Some surgical disorders are treated most commonly as an emergency because the involved organ does not usually present with chronic symptoms, e.g. acute appendicitis or acute Meckel's diverticulitis.

Emergency laparoscopy is nowadays increasingly used primarily to establish the diagnosis of acute right lower abdominal pain, but also to deal with conditions such as acute appendicitis and perforated ulcer. Overall, the benefit from interventional emergency laparoscopy for common general surgical emergencies (appendectomy, perforations, diverticulitis) can be substantial. Thus, in patients with a perforated duodenal ulcer, if the laparoscopy confirms that the perforation is well sealed, all that is needed is peritoneal lavage of the right paracolic gutter and pelvis and, perhaps, the insertion of a drain. Otherwise, the perforation is closed by suture or omental flap (depending on size) followed by lavage. It has to be stressed that the recovery of a patient from localized peritonitis caused, for example, by appendicitis/perforations is determined by the severity of the peritoneal infection and is not materially influenced by whether the appendix is removed through a grid-iron incision or by the laparoscopic approach. The argument that laparoscopic appendectomy is not justified because it does not reduce the hospital stay is illogical, especially as it significantly reduces the wound infection rate. Once the diagnosis of acute appendicitis is established by laparoscopy, the choice between open and laparoscopic appendectomy depends on the surgeon's preference and expertise. There have been 12 randomized trials comparing open with laparoscopic appendectomy. All have shown equivalent safety. The benefits for the laparoscopic approach registered by the majority of these studies are improved diagnosis of other causes of peritonitis (appendix inflamed secondarily as a result of disease elsewhere, e.g. perforated diverticulitis), reduced surgical site (wound) infection and earlier return to school and games in children. Laparoscopic interventions for acute small bowel obstruction are still only practised in a few centres, and there is no consensus on its benefits and risks. However, there is considerable reported experience on the benefits of laparoscopic treatment of the acute complication of colonic diverticular disease. Colonic obstruction by carcinoma is nowadays managed by stenting prior to laparoscopic or open resection.

Lack of clinical improvement in acute disease

Some patients have signs and symptoms that suggest an impending major complication. These patients are usually in hospital on conservative management or medical therapy with frequent observations and assessment of physical signs. Lack of progress with conservative management is an indication for *urgent (planned or scheduled)* intervention. In many instances, this decision is taken in consultation with the gastroenterologist and anaesthetist. A typical example is the acute exacerbation of ulcerative colitis with megacolon that does not settle with high steroid therapy.

Remedial

This term is reserved for operations on patients designed to correct:

- an iatrogenic complication
- persistent adverse symptoms that occur as a consequence of an operation that was otherwise successful in curing the patient of the disease
- deformity and restoration of the patient's body image.

By and large, these are specialized operations conducted by expert surgeons and the interest of the patients is best served by referral to tertiary referral centres. The repair of bile duct strictures following iatrogenic injury to the common bile duct during cholecystectomy is a case in point. The general surgeon should not attempt correction and is well advised to refer the patient to a hepatobiliary centre. There is good evidence that the best results in patients with postcholecystectomy bile duct strictures are obtained when the first corrective operation is conducted by a hepatobiliary surgeon experienced in dealing with this pathology (see Chapter 25).

Remedial surgery for persistent disabling symptoms is usually encountered in patients after surgery on the upper GI tract, e.g. vagotomy, gastric resections and antireflux surgery. These patients are very difficult to treat because they often have a multiplicity of symptoms, although, on careful history-taking, one is always able to determine the dominant symptom. Remedial surgery is aimed primarily at correcting this. In general, remedial surgery for adverse GI symptoms results in amelioration rather than cure, although the gain from surgery may be substantial, e.g. a patient with small stomach syndrome or severe dumping can eat, or a patient with severe intractable diarrhoea can lead a reasonable social life. The management of these patients entails careful selection, choice of the appropriate operation and expert dietetic management. Again, management in a tertiary referral centre gives the best results (see Chapter 23).

Surgery to deal with deformity is usually conducted after ablative operations for cancer, e.g. breast reconstruction after mastectomy, although primary reconstruction at the time of mastectomy is undertaken more frequently nowadays. Scar revision, usually following surgery on the face, neck and exposed extremities, is undertaken by plastic and reconstructive surgeons.

Ablative or functional

Operations can also be classified as ablative or functional. In the course of an ablative operation, the surgeon removes a block of tissue (surgical specimen). Such operations can be minor, e.g. removal of a subcutaneous swelling; moderate, e.g. cholecystectomy and partial thyroidectomy; or major, e.g. curative resections for cancer. It is axiomatic that all resected specimens must be submitted for histological examination. In cancer surgery, resections are classified as:

- *potentially curative*: assessment by surgeon at the end of the operation
- *definitely curative*: assessment after detailed histological examination: T, N and tumour-free margins
- *potentially non-curative*: after histological examination, any of serosal involvement, positive regional nodes, positive tumour margins (includes circumferential)
- *debulking*: reduction of tumour volume before either intraperitoneal chemotherapy or adjuvant chemotherapy
- *palliative*: (1) cancer is resected but there is residual disease; (2) bypass procedure for non-resectable cancer.

Obviously, for all potentially curative resections for cancer the surgeon must achieve a block resection with related regional lymph nodes up at least to level II and there must be no macroscopic evidence of residual disease (R0 resection).

In contrast to resections, functional operations are designed to correct a pathophysiological defect or abnormality, e.g. cervicodorsal sympathectomy for hyperhidrosis, cardiomyotomy for achalasia and fundoplication for pathological gastro-oesophageal reflux. As previously indicated, functional operations are best conducted through the minimal access approach as the access trauma (laparotomy or thoracotomy) constitutes the major component of the operative traumatic insult to the patient.

Prophylactic

These operations are only undertaken in patients at risk to prevent the development of serious disease, usually cancer, and thus increase life expectancy in the individual patient. Examples include prophylactic proctocolectomy in patients with familial polyposis coli and in patients with longstanding total ulcerative colitis on a surveillance programme. It is likely, with advances in molecular genetics and the identification of genes responsible for hereditary cancer, that more prophylactic operations including mastectomy will be undertaken in the future. To a large extent, bariatric surgery (surgery for morbid obesity) has a prophylactic intent as the weight loss not only improves the quality of life of the patient (immediate effect) but will reduce the risk of heart disease and disabling osteoarthritis/spondylosis, and increase the life expectancy and quality of life of these patients. As it also corrects maturity-onset diabetes, it is now referred to as bariatric and metabolic surgery (see Chapter 28).

Validated versus non-validated

Validated operations are those in established practice, the efficacy and outcome of which are known from past long-term usage or reported clinical studies/series. Non-validated operations are new procedures that have not been evaluated in terms of their efficacy and safety. The fact that an operation is carried out by the minimal access approach does not necessarily mean that it is a non-validated operation. Thus, for example, laparoscopic cholecystectomy, even when it was introduced, was a validated operation as the nature of the operation was unchanged from traditional open cholecystectomy. By contrast, laparoscopic hernia repair, when first introduced, was not a validated operation because both the anatomical approach and the nature of the repair differed from the traditional anterior inguinal canal approach. Subsequent clinical trials have established the merits and superiority of the posterior laparoscopic approach, which has largely replaced open repair of groin hernia.

It takes some time (years) to validate a new operation or interventional procedure. Heneage Ogilvie said of gastric operations that 'every gastric operation is good until it is found out'. In the ideal situation, new operations or interventions should initially be conducted in a few selected centres with strict external audit of the results. This preliminary assessment determines efficacy and safety, the documentation of which is essential before dissemination to other centres (after adequate training of staff).

There is a proficiency-gain curve for a new operation or intervention or surgical approach during which the surgeon's ability to perform the operation improves to the level of proficiency needed to obtain consistent good results with minimum morbidity. The problem is to ensure that surgeons go through this proficiency-gain curve without inflicting morbidity on the patients. This can only be achieved by supervised training, by which is meant assistance by an approved expert tutor until proficiency is reached. This approach ensures operative training with minimum morbidity and mortality.

Surgical exposure

The adequate exposure of the anatomical site of the operation is crucial to the efficient and safe execution of the procedure. Until recently, maximal exposure with a large enough incision was considered axiomatic. The advent of minimal access surgery (MAS) would, at first glance, appear to have changed this surgical principle, but in fact it has not.

The surgical exposure needed depends on the operation and the build of the patient. In practice, the options are:

- open surgical exposure
- laparoscopic exposure with positive pressure pneumoperitoneum
- laparoscopic exposure without gas insufflation – gasless (isobaric) laparoscopy
- hand-assisted laparoscopic surgery (HALS)
- thoracoscopic approach
- hand-assisted thoracic approach
- single port/incision laparoscopic surgery
- endoluminal surgery
- natural orifice surgery
- natural orifice transluminal endoscopic surgery (NOTES).

Open surgical exposure

Abdomen

Laparotomy or exposure of the peritoneal cavity is derived from the Greek *laparos,* meaning flank.

General considerations

Certain general considerations should be taken into account. In the first instance, the length of the incision depends on the thickness of the subcutaneous fat, which becomes a problem in the morbidly obese. These patients require large incisions and the subcutaneous fat should be split by distraction of the wound edges rather than cut, as this causes less bleeding from the large subcutaneous veins embedded in the fat (difficult to control by electrocoagulation). Second, females have a more commodious intraperitoneal space (including pelvis) than males who, especially when young and muscular, have tightly packed organs. The thicker musculoaponeurotic layer in males adds to the exposure problems and necessitates more strenuous retraction. In general, males have a deeper upper abdomen and, for this reason, access to the subdiaphragmatic space is more difficult. This becomes a problem in males with chronic obstructive airways disease and a barrel chest.

In addition to obesity, the type of build of the patient (asthenic, hypersthenic) is also important in the correct choice of incision. Thus, in patients with hypersthenic build and a long narrow subcostal angle, an upper transverse/oblique incision gives a very limited approach to the supracolic compartment. These patients are better served by a midline approach.

Incisions

Abdominal incisions may be longitudinal (midline, paramedian), transverse, oblique or a combination. Transverse incisions heal more successfully and are less painful than longitudinal but, to provide equal access, they need to be longer. The paramedian longitudinal approach is a poor choice incision. It used to be popular because of its trap-door design: the rectus muscle is interposed between the incisions on the anterior and posterior rectus sheath when the wound is closed. However, this had a bad result as the medial mobilization of the rectus muscle from the anterior and posterior rectus sheath inevitably devitalized sections of this muscle close to the tendinous intersections and, for this reason, resulted in partial atrophy of the muscle. If a longitudinal approach is preferred, the midline incision through the linea alba is a better proposition. The midline incision is advised when the diagnosis is not clear (exploratory laparotomy).

Both oblique and transverse abdominal incisions are muscle cutting to a greater or lesser extent but do not usually cause denervation problems, contrary to the longitudinal muscle-cutting rectus incision. The exception is the right subcostal (oblique) incision used commonly for open cholecystectomy, In which damage/division of the right subcostal nerve is almost inevitable and may cause subsequent neuralgic-type pain. The grid-iron incision (transverse or oblique) centred over McBurney's point (junction of the outer with the middle third of a line joining the anterior–superior iliac spine and umbilicus) is a muscle-splitting incision used routinely for open appendectomy. It may cause damage to the ileoinguinal nerve and thus weaken the musculoaponeurotic components of the inguinal canal, thereby predisposing to subsequent formation of an inguinal hernia on the ipsilateral side.

The bilateral subcostal incision used for wide exposure of the supracolic compartment of the abdomen is an example of a combination incision. Often, a third midline component is added between the two subcostal incisions and this combination (which may split the lower sternum) is referred to as the Mercedes incision. Other types of combination incision include the trap-door variety, in which one component is at right angles to the other. In the USA, a large curved transverse incision in the upper abdomen, which crosses the entire epigastric region, is often referred to as the 'sabre slash'.

Mini-laparotomy

The size in centimetres of a mini-laparotomy has never been defined and, to a large extent, the term is relative, i.e. the incision is smaller than the standard approach. On average, however, mini-laparotomy incisions are 7.0 cm in length. In practice, mini-laparotomy is used either as the only access for an operation, e.g. mini-cholecystectomy, or as part of laparoscopic-assisted surgery (see below).

Wound protection

In principle, wound protection is a sensible practice designed to prevent contamination by bacteria and thus wound infection and, in cancer patients, to reduce the incidence of implantation of exfoliated viable tumour cells and the subsequent development of tumour implantation deposits.

As almost all wound infections are endogenous in origin, protection of the operative wound from gross contamination is practised by some, either as plastic drapes or with antiseptic-soaked swabs (e.g. chlorhexidine). The disadvantage of plastic drapes is that their removal cannot be achieved without some contamination of the wound. To date, there has not been any firm evidence that the use of wound protectors significantly reduces the incidence of postoperative wound infection, which depends on (1) bacterial count of the wound edges at the end of the operation and (2) surgical technique (surgeon factor).

The exact incidence of tumour deposits in the wound following open surgery for cancer is not known. Current estimates are based on a large retrospective report on patients undergoing resections for colorectal cancer. This showed an incidence of 0.6% at 12 months, rising to 1% over the subsequent few years. Tumour deposits are related to the implantation of viable exfoliated tumour cells (spontaneous or induced by surgical manipulations). Lavage cytology studies have shown that GI tumours (stomach, colon) that have reached the serosa and pancreatic and ovarian cancer shed viable cancer cells spontaneously in 30–40% (in the absence of obvious peritoneal deposits). With the advent of MAS, concern arose from early reports of an increased incidence of tumour deposits in the access port wounds (port site deposits). In fact, the median reported incidence of port site deposits after laparoscopic surgery for potentially curative cancer is 1.2%, i.e. similar to open surgery; and again it appears to be surgeon related. Correct surgical technique based on oncological principles, selection of cases and wound protection during delivery of the specimen are crucial in the prevention of wound deposits.

Closure of abdominal incisions

Abdominal incisions are nowadays closed in a single layer (mass closure) in preference to layered closure, which had been popular in the past. Mass closure is quicker and gives consistently good results. Strong monofilament (1/0 gauge), as opposed to braided suture material, is preferred for closure of abdominal wounds for two reasons: (1) improved tissue slide and (2) less bacterial adherence. The monofilament material may be non-absorbable, e.g. Prolene, or biodegradable, e.g. PDS. Contrary to popular belief, nylon is biodegradable, as is silk. The use of large tension sutures to relieve tension on the suture line is favoured by some, especially when closure of the abdomen is difficult or prolonged postoperative ileus is predicted. If and when used, tension sutures should be tied loosely as otherwise they devascularize the wound.

Chest

The chest can be opened by (1) posterolateral thoracotomy, (2) prone thoracotomy – favoured in children, (3) anterolateral thoracotomy or (4) median sternotomy.

Posterolateral thoracotomy

This is the classic open access approach to the right or left chest cavity. The patient is placed on his or her side with the table split underneath the lower chest to widen the intercostal spaces. Appropriate supports for the arm, back and neck of the patient are essential. The incision is oblique and follows the line of the sixth rib. It should skirt a distance of 2.0 cm below the inferior angle of the scapula to avoid scapular adhesion after the operation. Posterolateral thoracotomy gives an excellent access but is accompanied by a high incidence of post-thoracotomy deafferentation pain. The right thoracic approach is favoured by most surgeons for subtotal oesophagectomy.

Correct positioning of the patient in a stable secure position is the responsibility of the surgeon. A double-lumen endotracheal tube (Carlen's or equivalent) is used to block and collapse the lung on the operating side (single-lung anaesthesia). In prolonged operations, this collapse is held partially responsible for some of the postoperative chest complications. Periodic reinflation for a few minutes intermittently during the course of the operation is practised by some to improve the cardiovascular condition of the patient during surgery and to reduce the risk of postoperative pulmonary collapse. Some patients are found at operation to be intolerant of single-lung anaesthesia, despite adequate lung volume estimates (spirometry) before the operation. In these cases, the surgeon has to operate with an inflated lung, using gentle compression to obtain the relevant access.

Anterolateral thoracotomy

The patient is in the supine position and the chest wall is open anteriorly on one or other side, usually along the fifth or sixth rib. A mini-anterolateral thoracotomy is used in video-assisted thoracoscopic surgery, e.g. lobar resection, pneumonectomy and Belsey mark IV procedure for reflux and hiatus hernia.

Median sternotomy

This provides access to the heart, great vessels and the trachea (central mediastinum) and to both sides of the chest. The skin incision is midline and the sternum is split by means of an electrical oscillating saw. Median sternotomy is the approach used for open cardiac surgery in combination with cardiopulmonary bypass (CPB), systemic hypothermia and cardioplegic arrest. The CPB systems use either a bubble or membrane oxygenator with roller or centrifugal pumps that provide a non-pulsatile flow. A heat exchanger is incorporated in the system for cooling and rewarming the patient. All CPB systems cause activation of the fibrinolytic and complement systems, and damage the blood cellular components. Nonetheless, CPB systems are quite safe for a period of up to 4–6 hours of use. Hypothermia (25–30°C) is used in most cases to decrease the patient's metabolic rate and tissue oxygen demands. Further selective cooling of the heart is achieved by topical cold solution within the pericardial sac and/or by infusion of cold (4°C) cardioplegia crystalloid solution into the coronary circulation.

Laparoscopic exposure

Positive-pressure pneumoperitoneum (capnoperitoneum)

In laparoscopic MAS, the standard exposure is through the creation of a positive-pressure pneumoperitoneum by an inert gas, traditionally CO_2; hence, the appropriate name should be positive-pressure capnoperitoneum. Other gases may be used, such as nitrous oxide, and this is preferred when laparoscopy is carried out for diagnostic purposes under local anaesthesia, as it appears to induce less pain.

The surgical exposure of the operative field by a positive-pressure capnoperitoneum (10–12 mmHg) is undoubtedly better than that achieved by open surgery, particularly for relatively inaccessible areas such as the subdiaphragmatic space and pelvis. Therefore, the surgical principle of maximal exposure of the relevant anatomy is not compromised. Indeed, the anatomy with the standard laparoscopes is magnified approximately two times. However, positive-pressure capnoperitoneal laparoscopic approach has some drawbacks:

- the charged couple device (CCD) camera used provides a tunnel view without any peripheral vision of the rest of the abdominal cavity
- the CO_2, which is kept at a set pressure (10–12 mmHg) by the automatic insufflator, creates convection currents that may be responsible for the spread of exfoliated tumour cells or bacteria
- the CO_2 dries and cools the patient's intra-abdominal contents unless heated and humidified by the insufflator
- the raised intra-abdominal pressure has adverse respiratory, cardiovascular, hormonal and metabolic effects.

There is evidence that a positive pressure capnoperitoneum above 12.0 mmHg reduces venous return and portal vein and splanchnic blood flow. Additionally, unless the gas is heated and humidified, the core body temperature of the patient drops significantly during long laparoscopic operations in addition to drying of the intestinal small bowel loops. Although with good anaesthesia the adverse haemodynamic changes can be tolerated by fit patients, the same cannot be said for patients with cardiorespiratory compromise. There is a clear correlation between the incidence and severity of adverse cardiovascular and respiratory complications and the level of positive pressure used for insufflation of the peritoneal cavity. Thus, the operative principle is to use the minimum pressure that gives adequate exposure. In children, this should rarely exceed 6.0 mmHg.

Gasless (isobaric) laparoscopy

Gasless laparoscopy is based on the use of technology, which elevates the anterior abdominal wall and thus creates the necessary work space. Its purpose is to abolish the adverse cardiovascular effects of the positive-pressure capnoperitoneum. The initial techniques consisted of either insertion of slings through the anterior abdominal wall around the midline and surrounding the falciform, or metal 'hooks' of varying shapes. Both are fixed to a chain that is then attached to an overlying gantry once sufficient lift is achieved (Figure 2.3). The anterior wall sling is simple and very useful. It permits conduct of a standard laparoscopic procedure (e.g. cholecystectomy, antireflux surgery) with the intra-abdominal pressure set at 4.0 mmHg. There is now good evidence from randomized studies that abdominal wall-lift devices (completely gasless or low-pressure pneumoperitoneum) substantially reduce adverse cardiovascular, hormonal and metabolic changes, accelerate recovery from anaesthesia without the problems of CO_2 narcosis and are accompanied by less postoperative pain.

However, the first generation of abdominal wall-lift devices had two major disadvantages: (1) tenting effect with reduced exposure and (2) trauma to the parietal peritoneum of the anterior abdominal wall. The space created by these lift devices is a conical space with the apex at the point of elevation by the retracting device, compared with the ovoid work space created by positive-pressure pneumoperitoneum. It is thus difficult to undertake complex operations with the gasless approach and the risk of iatrogenic injury during the dissection may be increased.

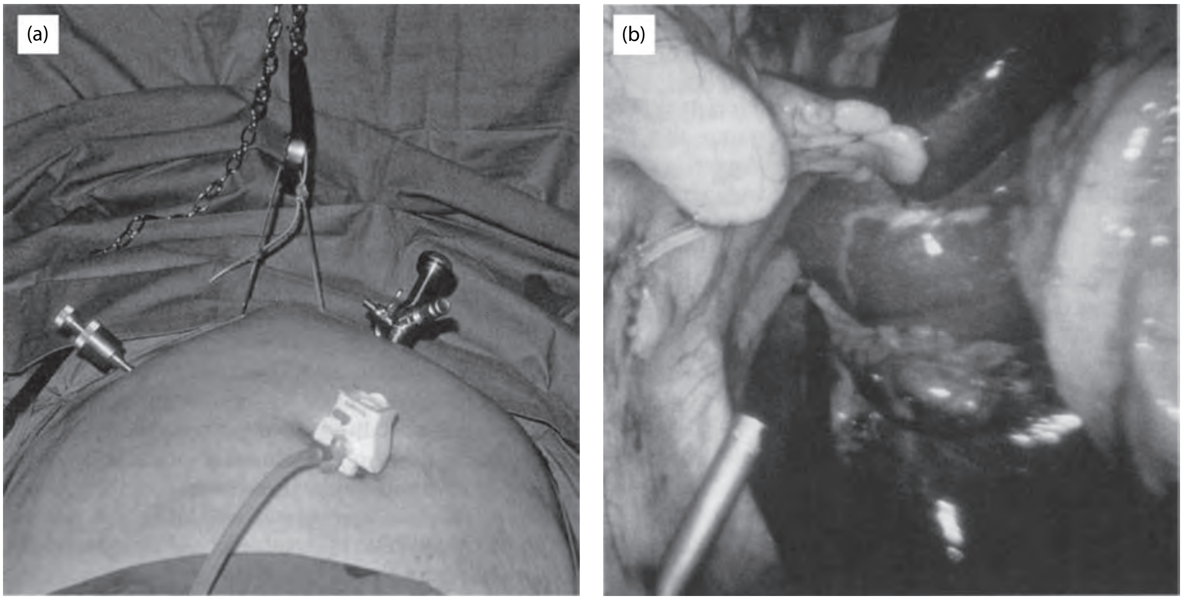


Figure 2.3 Right: surgeon operating from da Vinci console; left: robotic arms covered with sterile plastic drapes with end-effectors (instruments) introduced through parts inside the insufflated peritoneal cavity.

The second problem concerns the sustained compression/ischaemic trauma by the retractor on the parietal peritoneum of the anterior abdominal wall, especially during long procedures. This will encourage adhesion formation and, possibly, tumour implantation in operations for cancer.

The second generation of abdominal wall-lift devices have reduced but not abolished these problems by (1) providing extraperitoneal elevation and (2) minimizing the tenting effect. One example is the LaparoTensor (Figure 2.4), which is based on the insertion of two curvilinear needles in the subcutaneous plane of the anterior abdominal wall over the operative field. When inserted, the needles enclose a large circle open at one end. The butts of the needles projecting at one end beyond the skin are clamped together and to the lifting device. The shape of the needles has been designed by mathematical modelling of the abdominal wall to provide a better work space (the apex of the cone is widened), although this is still not as commodious as that provided by positive-pressure pneumoperitoneum. The gasless laparoscopic technique is seldom used in general surgery but the technique is favoured by gynaecologists for certain minor procedures such as laparoscopic myomectomy.

Extraperitoneal and retroperitoneal exposure

In this situation, there is no natural cavity, but one is created by finger and balloon distension of the loose areolar tissue surrounding an organ or anatomical region. Once the space is created, CO_2 insufflation through the optical port maintains the exposure. This approach is used for endoscopic adrenal surgery (as an alternative to the laparoscopic approach), lumbar sympathectomy, transabdominal preperitoneal groin hernia repair, pelvic node dissection for tumour staging of urological and pelvic tumours, etc. There is some evidence that the risk of CO_2 embolism is higher during retroperitoneal endoscopic than laparoscopic (intraperitoneal) surgery.

Combined exposure: laparoscopically assisted

There is no precise definition of laparoscopically assisted surgical operations. The endoscopic purist would regard an operation as laparoscopically assisted when the entire dissection is completed laparoscopically, and a mini-laparotomy is used at the end of the operation to deliver the specimen and perform the reconstruction of the gut, e.g. laparoscopic partial gastrectomy, laparoscopic right hemicolectomy and laparoscopic left colectomy or anterior resection. The ratio of the endoscopic to the external component of the operation can be varied and, in some instances, the combined approach is used to reduce the size of the abdominal incision for an open intervention. A pertinent example of this situation is anterior resection when, for whatever reason, the laparoscopic approach is contraindicated. In this situation, preliminary mobilization of the splenic flexure and adjacent transverse colon will permit an 'open' anterior resection through a transverse suprapubic (Pfannenstiel) or a lower midline incision. This combination provides 'the best of both options'.

There is no evidence that laparoscopically assisted operations are safer than the totally laparoscopic intervention in terms of the risk of iatrogenic injuries. The adoption of the combined approach, however, reduces the operating time and results in a better functional reconstruction of the GI tract. This is particularly the case after partial gastrectomy.

Hand-assisted laparoscopic surgery

There is a fine but important difference between laparoscopic-assisted and HALS. In laparoscopic-assisted surgery or GI operations, the procedure (dissection and excision) is carried out by the total laparoscopic approach. Then an appropriately sited abdominal incision (5–7 cm depending on specimen) is performed and, following insertion of a wound protector, the stomach or bowel containing the diseased segment/neoplasm

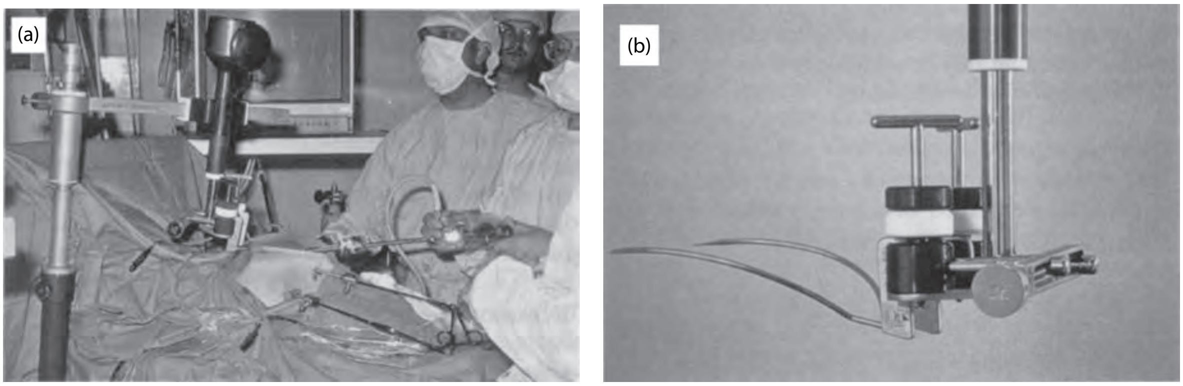


Figure 2.4 (a) LaparoTensor in use; (b) the subcutaneous needles provide a more uniform lift than intraperitoneal lift devices.

is extracted. After external resection of the lesion, continuity of the GI tract is restored by external anastomosis (suture or stapled) of the two ends. In HALS, a similar incision is placed at the start of the operation for the insertion of a hand port (various devices available) that allows the insertion of the non-dominant hand of the surgeon, or indeed, in some cases, the hand of the assistant, without loss of the positive capnoperitoneum. There is little doubt that the HALS approach facilitates the conduct of major and difficult operations for several reasons:

- restoration of the of the tactile feedback
- efficient and atraumatic hand retraction and exposure
- immediate control of haemorrhage.

There is good published evidence that the HALS approach does not reduce the benefits of MAS as, in comparative studies, the recovery is equivalent to that of the total laparoscopic approach. The HALS approach does reduce execution times for advanced operations, e.g. total R2 gastrectomy, subtotal oesophagectomy, pancreaticoduodenectomy and right hepatectomy. The HALS approach is essential for splenectomy for massive splenomegaly (see Chapter 26).

Thoracoscopic and video-assisted thoracic approach

The advantages of avoiding a large thoracotomy include accelerated recovery, especially in patients with compromised cardiorespiratory reserve, and less pain. The instruments and equipment used are similar to those used in laparoscopic surgery. There are several techniques for safe endoscopic access to the pleural cavity and the procedures can be carried out with (1) single-lung anaesthesia (endobronchial occlusion by a double-lumen tube) or (2) with bilateral pulmonary ventilation, when intrapleural CO_2 at a pressure of 6–8 mmHg is used to compress the lung. This level of positive intrathoracic pressure is well tolerated by most but not all patients.

Thoracoscopic approach

The complete thoracoscopic approach is now the standard treatment in many centres for pulmonary wedge biopsies, mediastinal node biopsy, for treating patients with spontaneous pneumothorax, for sympathetic denervation, mobilization of the intrathoracic oesophagus, for pulmonary segmentectomies and assessment/management of chest injuries. The approach can be anterior, posterolateral or posterior depending on the procedure and preference of the surgeon. The prone posterior jack-knife approach (Figure 2.5) is particularly useful for sympathetic denervation (cervicodorsal ganglionectomy and splanchnicectomy), for mediastinal node biopsy and for mobilization of the intrathoracic oesophagus. The telescope port is placed 2.5 cm below and in line with the inferior angle of the scapula with the elbows of the patient suspended along the edges of the operating table.

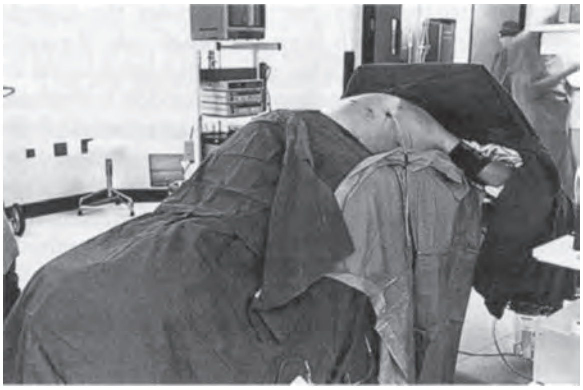


Figure 2.5 Prone posterior thoracoscopic approach. This allows access to both pleural cavities without changing the position of the patient. In addition, simple procedures can be carried out without single-lung anaesthesia.

Video-assisted thoracoscopic approach

In this instance, a 5.0 cm mini-thoracotomy (equivalent to mini-laparotomy) is made at the start of the procedure. Its exact location depends on the nature of the operation being performed. Video-assisted thoracoscopic approach is used for major interventions such as Belsey mark IV fundoplication, pulmonary lobectomies, pneumonectomy and coronary artery bypass surgery.

Endoluminal surgery

This refers to operations performed through operating flexible endoscopes or platforms and is well established in routine practice for submucosal resection of early gastric cancer using various techniques developed in Japan. Often saline or other fluid is injected submucosally to lift the mucosa containing the lesion, before it is resected.

Endoluminal surgery differs from NOTES as the integrity of the wall of the hollow organ (usually stomach, oesophagus and colon) is not intentionally breached and inadvertent perforation is a complication that needs immediate treatment; which at times can be performed endoscopically but often requires a laparoscopic or open operation. More recently dedicated systems have been introduced for the performance of endoluminal full-thickness fundoplication and the early reported results on their efficacy at achieving lasting control of gastro-oesophageal reflux have been promising but the reported experience and follow-up is too limited to enable a valid assessment on their ability to replace laparoscopic fundoplication. More recently, endoscopic myotomy 'POEM' has been shown to give equivalent results to laparoscopic cardiomyotomy for achalasia.

Natural orifice transluminal endoscopic surgery

NOTES is a novel technique utilizing the body's natural orifices to gain access to the peritoneal cavity for abdominal surgery. The goal is to perform intraperitoneal operations through a natural orifice using novel technology platforms based on modified operating flexible endoscopes. The perceived advantages include: (1) avoidance of abdominal incisions and thus painful wounds since no somatic nerves are involved and (2) the interventions can be carried out under sedation rather than general anaesthesia. The concept has generated new operating technologies by the flexible endoscopic industry, which is still evolving. The Natural Orifice Surgery Consortium for Assessment and Research was set up in the USA in 2005 as a joint initiative supported by the American Society for Gastrointestinal Endoscopy and the Society of American Gastrointestinal and Endoscopic Surgeons to oversee and guide through the development and introduction of NOTES into clinical practice. In Europe, a similar consortium was set up soon after by the European Association of Endoscopic Surgeons in association with industry for the same purpose. NOTES is undoubtedly a novel concept in human surgical intervention aimed at improved cosmesis (no visible scars) and reduction of postoperative pain; but, at the same time, it challenged some of the fundamental deeply entrenched dogmas in surgery and gastroenterology, including the important violation of asepsis as the operating system traverses a potentially infected environment before entering the peritoneal cavity, despite preparation of the hollow organ (stomach, colon, oesophagus, urinary bladder, etc.). It thus generated a series of experimental studies aimed at exclusion of the risk of contamination and infection of the peritoneal cavity and of techniques and technologies that ensure safe full thickness closure of the perforation of the hollow organ. Insofar as the stomach is concerned, these studies have confirmed that, with adequate preparation (gastric lavage with povidone–iodine solution and with intravenous antibiotic cover), the risks of infection into the peritoneal cavity are low and acceptable. Equally, the efficacy and safety of various technologies and techniques for safe closure of the gastric perforation have been confirmed. Even so, NOTES, despite the latest technology platforms, is difficult and requires two operators: one manipulating the platform and the other performing the procedure. Thus, despite the initial expectations, the uptake of NOTES by surgeons worldwide has been minimal and limited to a small number of transgastric NOTES operations, e.g. appendectomy, cholecystectomy, tubal ligation, replacement of percutaneous endoscopic gastrostomy, laparoscopic cancer staging and volume reduction of the gastric pouch after gastric bypass surgery for morbid obesity, and, although some of these interventions were performed with the pure NOTES approach, others have required insertion of abdominal laparoscopic ports (hybrid NOTES). The current view is that, unless robotic-based technologies are developed specifically for NOTES, this surgical approach is unlikely to become established in routine surgical practice.

Natural orifice surgery

This is related to NOTES but is less problematic as it poses fewer technical problems and infective risks. It is best exemplified by transvaginal interventions following the insertion of operating ports into the pouch of Douglas through the vaginal ports. To date, several hundred cases of vaginal cholecystectomy have been reported in humans with good results, but case selection is important. An interesting new development of natural orifice surgery is transanal endoluminal colonic surgery being developed in Europe by Dr Antonio Lacy. The procedure starts with insertion of the TEM (transanal endoscopic microsurgery; Wolfe) or TEO (transanal endoscopic operation; Storz) operating proctoscope. The rectal wall is divided with electrosurgery just above the dentate line through the operating proctoscope. The distal end is then closed by a continuous suture leaving the rectum and anal canal as a conduit to the pelvic peritoneal cavity. Both sigmoid and extended left hemicolectomy have to date been performed but the reported experience is limited. The specimen is extracted per annum and is followed by a stapled anastomosis guided by a single laparoscopic port for the camera.

Single port/single incision laparoscopic surgery

The trend towards reduction of traumatic insult to patients has continued in the quest for reduction of the invasiveness of surgical interventions. In this development single-port laparoscopic surgery (SPLS; also known by a variety of other names, e.g. SILS for single incision laparoscopic surgery, reduced port surgery) has experienced the greatest uptake by surgeons, far exceeding NOTES. The concept of single-port endoscopic surgery is not new as it was used initially by Dr Raimund Wittmoser for single-port thoracoscopic surgery on the sympathetic and vagal nerves in the 1970s. One-wound laparoscopic cholecystectomy was first reported in 1997. Resurgence in the interest and uptake of SPLS occurred with the advent of NOTES in 2005 with substantial advances in port technology, novel retraction systems and related instrumentation.

There is no doubt that SPLS imposes more ergonomic restraint than traditional multiport MAS: (1) insufficient triangulation, which limits the manipulation angle (ideal = 60° triangle) between two working instruments held in the dominant and non-dominant hands; (2) internal and external collisions between the instruments and optic is more problematic when standard straight laparoscopic instruments and a 0° (forward viewing) telescope are used; (3) problems with retraction and exposure; (4) enforced approximation of the surgeon's hands by the single port or incision; all of which add to the difficulty in the execution of the operation.

Thus, the traditional methods used in conventional multiport laparoscopic surgery have to be adapted for SPLS. Often, additional ports (reduced port surgery) are used to insert instruments specifically for exposure and retraction.

Technologies for single-port laparoscopic surgery

Considerable effort in research and development has been made and techniques devised in the effort to restore, at least to some degree, the triangulation and a reasonable operating angle. The majority of single-port devices for SPLS are disposable but two are reusable: the EndoCone and the X-port (Storz, Tuttingen, Germany). The EndoCone has a detachable bulk head (cap) that contains six latex valve inlets admitting use of 5 mm instruments and two central inlets that allow the insertion of large instruments including a stapler (Figure 2.6). The surgeon is able to use three instruments and an optic at any one time during the course of the operation. This EndoCone has been successfully used for laparoscopic right and left hemicolectomy splenectomy, and nephrectomy and sleeve gastrectomy.

The commonly used disposable single-port devices are the SILS port (Covidien), the TriPort (Olympus), the GelPOINT (Applied Medical) and the AirSeal (SurgiQuest). The SILS port is a blue flexible soft-foam port. It is designed to be inserted through a single 15–25 mm periumbilical incision. The device has three access channels that can accommodate three 5 mm cannulas or two 5 mm cannulas and one 12 mm cannula. The main drawback of the SILS port is that it is associated with difficult insertion, lack of abdominal wall adjustability, and insufficient maintenance of capnoperitoneum during the operation. The TriPort Access System has three instrument channels that can accommodate two 5 mm instruments and one 12 mm instrument and the device can be inserted through a single incision of 12–25 mm. It has one distal ring and one proximal ring. The GelPOINT system consist of the Alexis wound protector/retractor, GelSeal cap and 5 mm self-retaining trocars. The critical disadvantage of the GelPOINT is that it only has 5 mm self-retaining trocars, which limit the usage of larger instruments such as a stapler.

Instrumentation and optics for single-port laparoscopic surgery

The curved coaxial instruments modified with proximally deviating handles (Figure 2.6a) are able to restore useful triangulation and overcome the problem of enforced approximation of the surgeon's hands. Another useful development relates to instruments of variable length, which enable placement and manipulation at different levels and planes and increase the external working space close to the port, thereby avoiding external collision. A more advanced development for SPLS concerns articulated instruments that can allow work

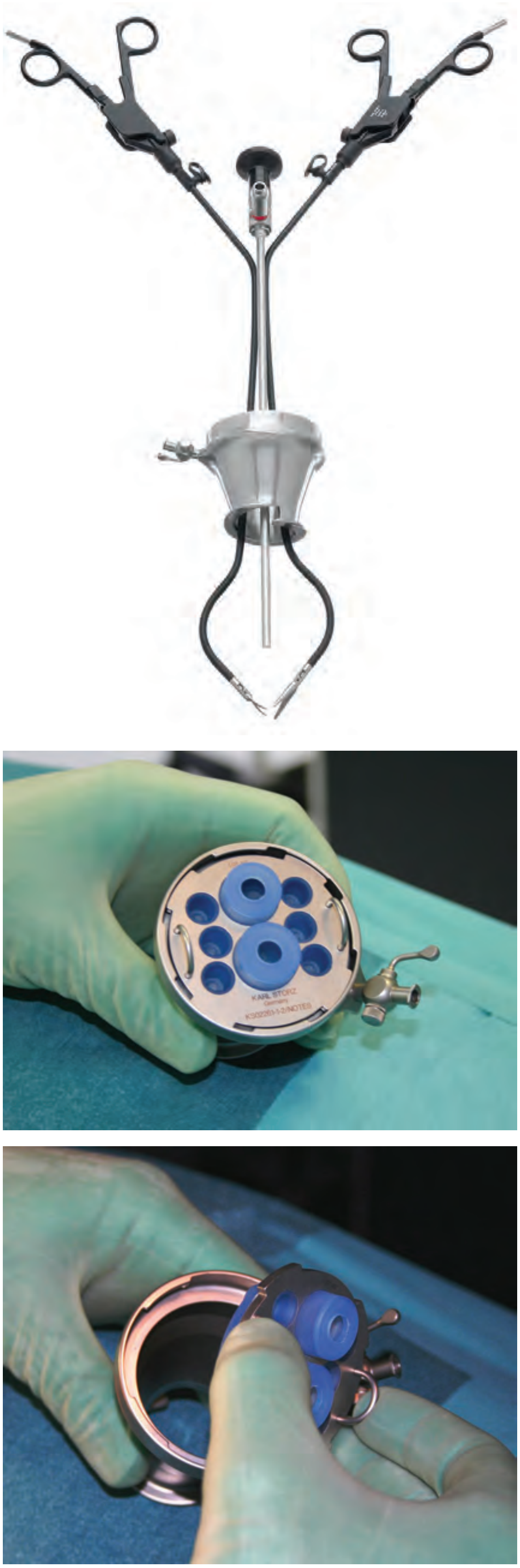


Figure 2.6 (a) Dundee EndoCone, which is self-retaining after insertion using a screwing action with proximally deviating handles (to separate the surgeon's hands) and distal curved instruments (to restore useful triangulation). (b) EndoCone fitted with a cap containing the ports. (c) The cap is detachable for extraction of small specimens through the cone, thus acting as a wound protector.

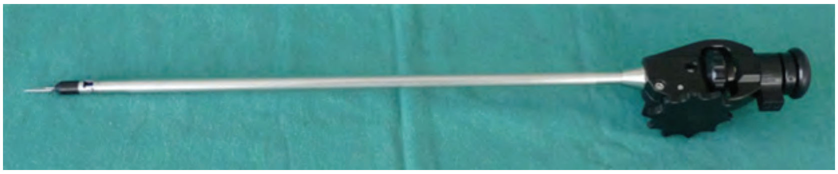


Figure 2.7 Dares hand-held manipulator with seven degrees of freedom. The instrument has a flexing and rotating functional tip.

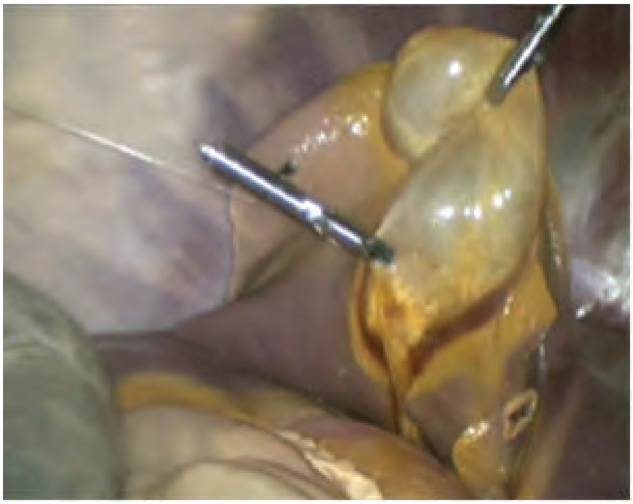


Figure 2.8 Dundee anchored and retraction tether system (DARTS). This uses small, tethered self-grasping anchors (instead of sutures) introduced through a modified Veress needle-type introducer. Image shows exposure of the gall bladder.

in the operative field despite a straight proximal segment. Aside from reducing external collision of instruments, they help to restore triangulation at the tips of the instruments. More advanced systems include the hand-held manipulations with six degrees of freedom (DOFs). These manipulators enable controlled adjustable bending of the distal segment of the instrument and rotation of the jaws of the instrument, thereby considerably facilitating intracorporeal suturing during SPLS operations (Figure 2.7).

New optics for SPLS include the Olympus EndoEye (Olympus), which has a distal chip and yields more working space close to the access port. Also, some surgeons prefer to use flexible-tip endoscopes to reduce the 'moving together' effect between the scope and instruments.

Retraction systems for single-port laparoscopic surgery

Suspension sutures are often used for retraction and exposure of Calot's triangle for SPLS cholecystectomy. Dedicated systems such as Endo-retractor have been used clinically. A new device, known as DARTS, is currently under evaluation in preclinical studies (Figure 2.8). The DARTS retraction system uses small atraumatic self-grasping tethered 1.5 mm clamps introduced through a modified Veress-type placement system that function in the same way as suspension sutures but avoid the difficulty of suture placement.

In human cadaveric tests, the ability of DARTS to provide successful retraction and surgical exposure of the cystohepatic angle through a single incision without use of extra ports has been demonstrated.

Dissection and fascial planes

'There is no such thing as a surgical postoperative complication – all are enacted on the operating table.' This quotation is attributed to a master surgeon from Liverpool. It is correct insofar as technical complications are concerned, but ignores the systemic factors that may complicate surgical intervention. Nonetheless, it stresses, by exaggeration, the importance of technical execution (task quality) to the clinical outcome. Whatever the complexity of the pathological anatomy and density of adhesions, the operation progressed at a steady pace and the tissue planes seemed to open up as he dissected. When watched, he appeared to be slow but, in fact, the operations were completed fairly quickly. When once asked to explain his technique, his reply was 'this is how I have always done it'. On reflection, the operative technique of this and other master surgeons had the following characteristics:

- immediate grasp in three dimensions of the abnormal anatomy of the case (spatial perception)
- appreciation of tissue planes
- economy of movements with purposeful manipulations
- operative choreography
- upper limb co-ordination and manual dexterity
- gentle but purposeful touch
- strict adherence to the sequential steps of an operation.

This surgeon had reached the stage when he rarely committed any execution errors. He had developed these skills over many years to an extent that his brain was conditioned to execute 'macro-operative programs'. He operated on autopilot and, for this reason, could not specify the components of his surgical operative skill. These are the skills that every surgical trainee should strive by constant effort to improve and to hone his or her operative manipulations. Some of these skills are worth considering in some detail.

Touch

The essence of good dissection is the application of the right amount of stretch on the tissue with the assisting hand before division/suture, etc. of the tissue with the active hand. The correct amount of stretch on the tissue (without trauma) is entirely dependent on the tactile force feedback that the surgeon has developed from the innate level with which he or she was born. It is the level of appreciation by the surgeon of this sensory feedback to the hand that distinguishes the surgeon with the gentle touch from the rough one. Research into psychomotor skills has shown that left-handed individuals are by nature more gentle than right-handed individuals, although the latter are more precise and commit fewer errors (swings and roundabouts). A gentle touch must be accompanied by accuracy and decisive manipulations by the active hand. Tentative active manipulations, e.g. scratching with the scissors or inadequate needle passage, should be avoided.

Identification of tissue planes

Normal tissue planes are made up of condensations of areolar tissue and are relatively avascular, permitting separation with ease and without significant blood loss if the surgeon proceeds in the correct plane. Otherwise, bleeding, difficulty and resistance to separation are encountered. These are indications that the dissection has deviated to the incorrect plane. These normal fascial planes are altered by previous surgery and inflammation: the fibrotic component is increased and this is most dense close to the visceral side, becoming less so as the parieties and retroperitoneum are reached. The transition between the dense fibrous layers and the outer areolar tissue forms a distinctive 'white line'. This indicates the correct plane of dissection.

Economy of movement

All movements made by the surgeon in performing tasks must be deliberate and purposeful, and achieve the desired objective. Unproductive movements not only waste time but also increase the risk of error. Another aspect of this skill requires that one step of an operation be completed before proceeding to the next. Thus, flitting from one step to another is counterproductive to the progress of an operation.

Choreography

Each operation is made up of a number of steps which, in turn, are composed of a series of sequential tasks. Choreography implies a coordinated performance of the various component tasks of a step such that the execution of one task facilitates the performance of the subsequent one. In this fashion, the step is executed smoothly and efficiently. Errors are of two kinds: intrastep (execution) or interstep (procedural). An intrastep or execution error means that the task is executed poorly, e.g. a needle swivels during passage through the tissue, or scissors cut the wrong layer. By contrast, a procedural or interstep error indicates missing a step or the wrong order of steps, e.g. failure to clamp or coagulate an arteriole before dividing it. In essence, choreography signifies that a surgeon has an orderly checklist of the steps of an operation in mind and, in executing the procedure, he or she follows this rigidly.

Types of dissection

Surgical dissection can be either blunt, i.e. the tissue planes are teased apart, or sharp, when a cutting device is used.

Blunt dissection

When used judiciously, blunt dissection is effective and safe. It requires the presence of loose areolar tissue that allows easy separation as a shear force is applied by hand or using a gauze swab or a pledget (peanut) held in a forceps (or long, secure grasper in endoscopic surgery). It relies for its safety on tactile feedback to the surgeon's hand and, unless this is refined, excessive shear force can be inadvertently applied so that tissues are torn and vessels avulsed. Blunt dissection is very useful in anatomically crowded areas or when the correct anatomy cannot be identified, especially if important structures, the integrity of which must be preserved, are known to be present within the block of tissue.

Sharp dissection

Sharp dissection, which involves division of tissue and separation of tissue planes, is carried out by mechanical (scissors, scalpel) or energized [high-frequency (HF) electrosurgery, ultrasonic devices, etc.] equipment. Irrespective of the equipment used, sharp dissection is a two-handed procedure that involves the application of the correct amount of tension by the assisting hand (often with atraumatic dissecting forceps, but sometimes directly with the hand over a gauze swab), while the division/separation is effected by the active hand. Clean separation and cleavage is as much dependent on the 'cutting instrument' as on the amount of stretch applied to the tissue.

HF electrosurgery is the most commonly used form of energized sharp dissection but, during the 1990s, vibrating (ultrasonic) equipment, water-jet and laser devices were introduced both in open and in endoscopic surgery. The physical principles governing the safe use of these forms of energized dissection are discussed in Surgical technology.

Adhesions

Adhesions result from previous inflammation or operative interventions within serous cavities. The reported incidence of adhesions after open general abdominal operations ranges from 67% to 93%, although there is great variation in the formation of adhesions between individuals. Adhesions are also much less common after laparoscopic surgery. The adverse consequences of adhesions are:

- intestinal obstruction
- difficult reoperative surgery
- increased risk of bowel damage during abdominal surgery
- chronic pain
- infertility in females
- increased healthcare costs.

The division of adhesions is an integral part of abdominal surgery. Adhesiolysis may be necessary to obtain exposure or because the adhesions are responsible for the primary pathology, i.e. mechanical small bowel obstruction.

Adhesiolysis

Aside from causing symptoms and intestinal obstruction, adhesions influence the surgical access used and the level of technical difficulty of an operation, whether this is open or endoscopic. Undoubtedly, adhesions increase the morbidity from iatrogenic injury in laparoscopic surgery, especially during the creation of the pneumoperitoneum, when adherent bowel can be impaled by the Veress needle or, worse, by the optical trocar cannula. There are now safe techniques for the creation of pneumoperitoneum in patients with adhesions from previous surgery. In open surgery, the technique for safe adhesiolysis relies on:

- establishing the anatomy of the adhesions
- putting the adhesions on the stretch (usually with the flat of the assisting hand over a swab), to identify the white line; this marks the insertion of the adhesion, which is completely avascular and can thus be divided by scissors without bleeding
- separating adhesions over a wide front; otherwise, the surgeon 'digs himself into a deep hole', with poor exposure and risk of damage to bowel loops.

Adhesions must not be separated by energized dissection and should not be clamped and divided between ligatures. This obscures the tissue planes and promotes further adhesions.

Tissue approximation

Accurate tissue approximation is essential for operative repair of defects and execution of safe anastomosis. Aside from gentle handling of tissues and careful dissection, the approximation must be achieved without tension and without compromise of the integrity of the blood supply essential to the healing process. The approximation itself can be performed by:

- sutures
- staples or clips or tacks
- glues.

Approximation by suturing

Approximation of edges involved in the anastomoses of hollow viscera, closure of wounds and repair of defects by interrupted or continuous suturing is an acquired craft based on established surgical principles. The key principles involved are:

- Preservation of the blood supply.
- Approximation of the edges without tension; this applies equally to repair of hernias and to intestinal anastomoses that should 'sit' rather than 'stand'.
- Meticulous technique with attention to detail and execution in the same way every time.
- Correct suture spacing and suture bites. It should be remembered that wounds heal despite sutures, which tend to devascularize the wound edges. Thus, especially in GI anastomosis, the balance between creating a seal and devascularization is fine. In this respect, crowding the anastomotic line with too many sutures predisposes to failure of the anastomosis. As a general rule, the space between the individual sutures should approximate to the size of the individual suture bites.
- Selection of the correct suture materials.

Continuous and interrupted suturing

Although in some instances either continuous or interrupted suturing can be used with equally good results, this is not always the case and certain specific anastomoses require the use of one and not the other of these two techniques. Continuous suturing is quicker than interrupted suturing and results in a more leak-proof and haemostatic join. Hence, it is the technique deployed in vascular anastomoses. The 'purse-string' effect inherent to continuous suturing is minimized by correct suture spacing (the greater the distance between suture bites, the more pronounced the purse-string effect) and by the use of separate sutures for the posterior and anterior suture walls of the anastomosis. Even so, such an anastomosis cannot enlarge and this is an important consideration in children, in whom an anastomosis has to allow for growth. Continuous suturing is inadvisable in the anastomosis of small tubular structures and is difficult in deep cavity work.

Interrupted suturing is slower but allows more accurate suture placement. It is the technique used routinely in microsurgical work. Although less waterproof than the continuous technique, interrupted suturing is preferred by most surgeons for oesophageal, small bowel and colonic/rectal anastomoses. The sutures are placed on the back wall of the anastomosis before they are tied and the procedure is then repeated for the anterior wall. This technique ensures correct suture placement and precise coaptation of the edges. Interrupted suturing is less likely to devascularize GI anastomoses. This is an important consideration in high-risk areas such as oesophageal and colonic anastomoses. Both interrupted suturing and stapling (see below) are accompanied by a higher incidence of postoperative suture line bleeding from anastomoses involving the stomach, particularly gastrojejunostomy.

Ligatures and suture materials

Ligatures, as opposed to sutures, are lengths of biocompatible thread used to tie structures such as blood vessels. The vast majority of sutures used in surgery nowadays are atraumatic, i.e. the needle is attached to the material by either swaging or crimping. In swaging, the rear of the needle is flattened and then folded over as a tunnel to hold the suture. Crimping, which is used for larger needles, entails drilling a canal at the rear of the needle (by laser) to accommodate the thread, which is then crimped in place. Crimping provides a more secure join between thread and needle. Swaging is used for fine atraumatic and detachable sutures (controlled release or pop-off), in which the attachment is intentionally set to a minimum strength, so that after passage through the tissue the needle can be easily pulled off. Controlled release atraumatic sutures are invaluable for interrupted suturing in deep cavity work. They should not be used for continuous suturing.

The attachment of the thread to the needle by either technique creates a problem because the diameter of the needle at the junction is always larger than the suture. In practice, this ratio averages 2:1 for large and 3:1 for small sutures. The adverse effect of this step is that the suture lies loosely in the tissue, as the tunnel made by the needle is at least twice the size of the suture. Atraumatic sutures made from expanded polytetrafluoroethylene (Gortex) provide an exception to this, since the compressible sponge-like nature of this material results in a ratio of 1:0.8, i.e. the suture fits snugly in the tissues, improving the seal.

All materials used, whether natural or synthetic, meet certain mechanical, biological and handling requirements.

The important mechanical characteristics are tensile strength, yield point, elasticity and breaking strength retention (BSR). A high tensile (breaking) strength is always a good attribute because it permits the use of a finer suture. In terms of the bulk of the suture material in contact with the tissue, a 2/0 suture has a circumference that is double that of the 3/0 equivalent; hence, the inflammatory response is higher and the risk of infection greater. The tensile strength is also important in relation to knotting, since the strength of a perfect knot is, at most, only 50% of the tensile strength of untied ligature or suture material from which it is made. The BSR is an important mechanical characteristic. It is a reflection of the rate and extent of biodegradability of the suture following immersion in fluids (*in vitro* BSR) or implantation in animal tissues (*in vivo* BSR). When a suture is subjected to progressive distraction in a tensiometer, a certain point is reached when the material starts to deform and stretch permanently before it breaks if the distraction force is continued beyond this yield point. Young's modulus of elasticity (which is measured from the start of the distraction force to the yield point) is an expression of the resistance to elongation of the suture before it breaks. The higher the modulus, the less elastic the suture. A measure of elasticity is a good attribute as it enables the material to undergo elastic deformation when tied, thereby increasing knot security. By contrast, excessive elasticity is a nuisance as the material lengthens and distorts (crinkles) during knotting. The ideal combination is a suture with a high tensile strength, good BSR and a low to medium modulus of elasticity. The biological characteristics depend on the composition, structure and surface charge of the material. They include biodegradability, reactivity (extent of inflammatory reaction), tissue incorporation (tissue growth into the interstitial microstructure of the material), effects of infection and site-specific adverse effects, e.g. deposition of calcium on non-absorbable material in the urinary tract and calcium bilirubinate in the biliary tract.

The surgical/handling characteristics relate to the ease of handling, slide through the tissue, knot run-down and knot security. The friction coefficient of a suture determines the ease of slide through the tissue, the knot run-down and the resistance to knot slip. Monofilament sutures slide more easily through tissues (and are thus less traumatic) than their braided equivalents. In the case of synthetic sutures, this disadvantage of braided sutures is largely, but not completely, overcome by coating the braid with a lubricant copolymer such as polybutylene. A low-friction suture exhibits excellent tissue slide and good knot run-down at the expense of an increased tendency to spilling. Thus, knots fashioned from these sutures require several throws (minimum of five) as opposed to the traditional three. The visibility of a suture is an important yet overlooked property. Only a limited range of dyes is approved for dyeing sutures and ligatures (violet, black, shades of blue–green). The visibility of a suture helps considerably in the execution of correct suturing and is an important characteristic in laparoscopic work, in which blue–green is the most appropriate. Transparent (undyed) biodegradable monofilament sutures are difficult to use, especially in the smaller gauges, but are preferred by plastic surgeons for subcuticular closure as the dyed sutures may lead to tattooing when the dye leaches out as the material degrades.

Table 2.1 Size of suture/ligature material: empirical gauge versus metric systems

Empirical gauge	Metric size (mm)
10/0	0.02
9/0	0.03
8/0	0.05
7/0	0.07
6/0	0.1
5/0	0.15
4/0	0.20
3/0	0.25
2/0	0.35
0	0.4
1	0.45
2	0.55

Traditionally, the size (diameter) of sutures and ligatures is expressed by the empirical gauge system. The equivalent metric size (diameter in tenths of a millimetre) is shown in Table 2.1. The gauge range used in general surgery is from 1 to 5/0 (0.45–0.15 mm).

Natural materials

These include catgut (plain, chromic, alcohol packaged, dry), silk, linen, cotton and collagen. In most countries catgut (which despite its name is made from bovine collagen) is banned because of the potential risk of bovine spongiform encephalopathy. From a physical standpoint, all natural suture materials can be considered as twines of fibrous polymers. As a group, they have excellent handling characteristics and knotting properties. However, they excite a prominent inflammatory response within the tissues and, moreover, are easily colonized by bacteria, which invade the interstices of the twine and are then less susceptible to antibiotics.

All of the natural materials (including silk) are biodegraded by proteolytic digestion. The degradation is rapid with catgut and slow with silk and cotton. Another property of the natural materials is that they absorb water (tissue fluid) and thus swell following implantation. This swelling increases the knot security and is probably the reason why black silk is still popular with some surgeons. The BSR of plain catgut (derived from bovine intestine) is poor, and this material loses 50% of its tensile strength within a few days of implantation. Chromic catgut has twice the BSR of plain catgut and is thus preferable, although it is nowadays banned in most countries.

Synthetic non-absorbable materials

These *hydrocarbon polymers* are all derived from coal and oil. They are hydrophobic and therefore do not swell following implantation. Although generally regarded as non-biodegradable, some, such as polyamide (nylon), degrade in the tissues slowly with time. The group characteristics include flexibility, high tensile strength, resistance to creep and relative inertness. The main disadvantage of these hydrocarbon polymers in the monofilament form is their poor knot-retaining quality, with a tendency to knot spilling. The development of multifilamentous copolymers (by braiding of multifilaments or by looping from a single filament) has solved this problem.

Polypropylene (olefin) is marketed as Prolene (Ethicon) and Surgipro (USSC). It is a monofilament material with a low friction coefficient (slides well through the tissues), excites a minimal tissue reaction and handles better than nylon. It is, however, elastic and can fracture. Knots tend to spill unless tied and drawn properly. Five throws are recommended for security. Polypropylene is used extensively in general surgery (e.g. mass closure of the abdomen) and for vascular anastomoses. Subcutaneous knots must be buried as otherwise localized tenderness/sinus formation may occur.

Polyamide (nylon) is available as monofilament (Dermalon, Ethilon, Monosot) and braided (Nuralon). Polyamide combines strength and elasticity but is slowly biodegradable, retaining two-thirds of its strength for up to 6 months. In the monofilament form, it is a difficult material to handle and does not tie well, with a tendency for the knot to spill. Despite this, it is popular with many surgeons for mass closure of laparotomy wounds. The braided form (Nuralon) handles and ties well and is used extensively for single-layer GI anastomoses by many surgeons.

Polyester (Dacron) is the strongest of the non-absorbable polymers and has the best mechanical and handling properties. It knots and handles very well. Polyester is available as the uncoated braid (Mersilene) and as a braided suture coated with polytetrafluoroethylene (Ethibond, Surgidac).

Expanded polytetrafluoroethylene (Gortex) is an excellent suture and excites the least inflammatory reaction of all biomaterials used in surgical practice. Its microporous structure allows tissue ingrowth. Expanded polytetrafluoroethylene handles extremely well, although the knots tend to spill unless locked tightly. A minimum of four throws is recommended for knot security.

Synthetic absorbable materials

These are much less reactive than catgut, i.e. cause less inflammatory reaction after implantation. They are degraded slowly by hydrolysis rather than proteolytic digestion and exhibit a BSR of several weeks. The degradation by hydrolysis is important as this process releases hydrogen ions, thereby lowering the pH in the vicinity of the wound. This is thought to reduce bacterial colonization and infection. Synthetic absorbable suture materials can be monofilament or braided. The latter are available as coated and uncoated braids (Table 2.2).

Table 2.2 Synthetic absorbable suture materials

Monofilamen	Braided
PDS (polydioxanone)	Braided polydioxanone
Maxon (polyglyconate)	Dexon (coated polyglycolic acid)
Monocryl (poliglecaprone)	Vicryl (coated polyglactin)
Biosyn (glycomor)	Polysorb (coated lactomer copolymer)

Needles

Surgical needles are essentially penetrating devices designed to pass sutures through tissue with minimum trauma. As mentioned previously, most sutures are nowadays attached to the needle by crimping or swaging (atraumatic sutures) and the French eyelet needle is rarely used. The important characteristics of surgical needles are shape, size, tip and configuration of the shaft (body).

Shape

Needles can be straight, curved, compound curved and customized. The large, straight Keith needle is still popular for hand suturing and small, straight needles are used by some surgeons for laparoscopic suturing. The straight needle is not ideal for suturing, however, because it does not allow the smooth wrist-scooping movement. In addition, the straight needle results in a 'cutting' effect as the tissue is stretched on the shaft when the needle is picked and pulled out at the exit point. Curved needles do not have these problems. They are based on various sections of a circle (1/8, 1/4, 3/8, 1/2 and 5/8). The most commonly used curved needles in general surgery are the 1/2 and 3/8 circle configuration. The 1/2 circle needle is ideal for end-to-side anastomosis, whereas the 3/8 circle is better for tissue plane approximation and end-to-end anastomosis. The compound curved needles have different profiles at the tip and body and are exemplified by the J-shaped needle, fishhook needle and the endoski needle used in MAS (Figure 2.9).

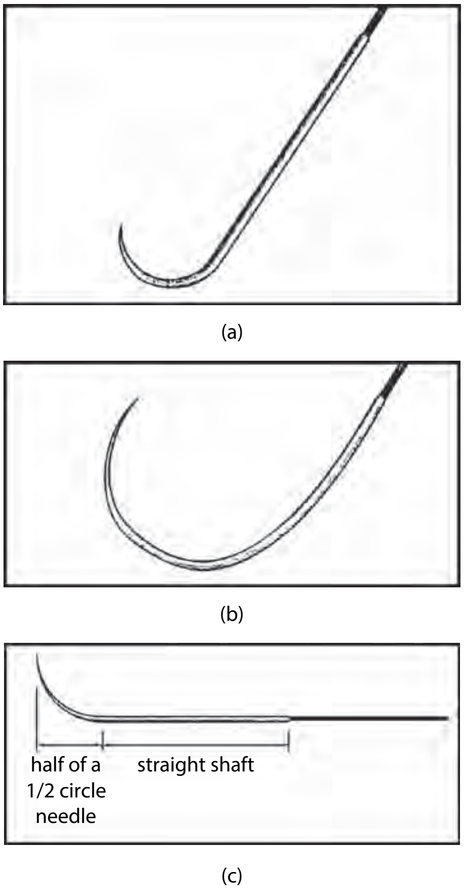


Figure 2.9 Examples of compound curved needles: (a) J-shaped; (b) fishhook; (c) endoski.

Size

This relates to the length and diameter of the needle. The length chosen is dictated by the distance between the entry and exit bites and by the length of the needle that needs to emerge from the exit point so that it can be picked up by the needle driver without damage to the tip. The gauge (diameter) chosen depends on the resistance of the tissue and its texture. In atraumatic sutures, the needle is always thicker than the suture and the ratio of the two is seldom below 2:1. The ratios of needle diameter to suture diameter for atraumatic sutures used in general surgery are shown in Table 2.3.

Tip

The sharpness of the tip and the tapering ratio determine the penetrating characteristics of the needle. The tapering ratio describes the extent of the taper from the maximal cross-section of the shaft (shoulder) to the needle tip and is calculated from:

$$\text{Tapering ratio} = \frac{\text{Distance between shoulder and tip}}{\text{Diameter of the shaft}}$$

The longer the taper, the more the needle penetrates atraumatically. Sharp needles with a long tapering ratio are used for suturing delicate tissues or those softened by inflammatory oedema. Sharp needles, however, increase the possibility of injury to the surgeons and operating staff. This is an important consideration in infected cases. Closure of the abdominal wound with blunt needles is advisable in these cases.

The tip of the needle is either conical, with a round cross-section (tapered), or cutting, in which the shape is pyramidal and the cross-section triangular, with either a cutting inside edge (conventional) or a cutting outside edge (reverse cutting), or both cutting outside and inner edges (cutting tip). Cutting needles are used for penetrating tough tissues, e.g. skin and cornea.

Body

The shaft or body of the needle is the longest section and extends from the shoulder to the join with the suture. Its cross-sectional shape determines: (1) the extent of trauma to the tissues during needle passage and (2) security of hold by the jaws of the needle driver (preventing deflection and swivel). The least traumatic configuration is the round-bodied needle, which should always be used for suturing delicate tissue, but this needle offers the least secure grip. The oval needle is fairly atraumatic and provides a better grip with fewer tendencies to swivel. It is preferable to the round-bodied variety in laparoscopic work. Various other shaft shapes are available for specialized work: ribbed, flattened hexagonal, triangular, etc.

Table 2.3 Atraumatic suture sizes and needle–suture ratios used in general surgery

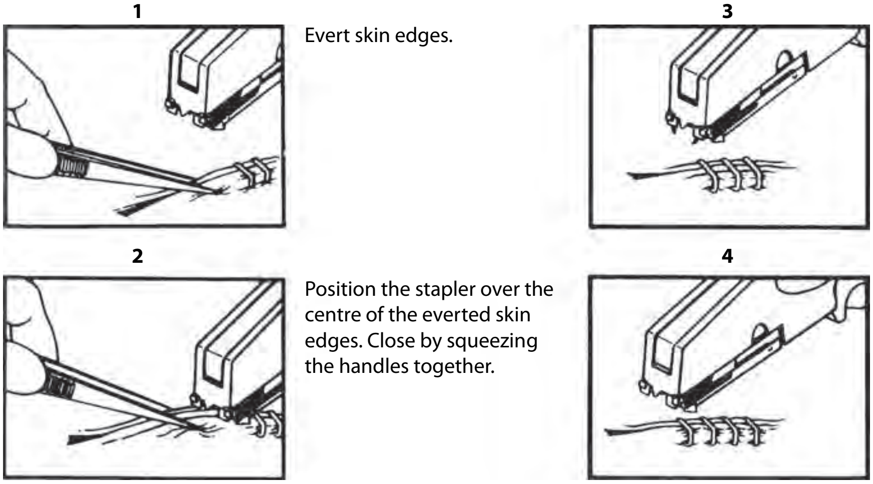
USP size	Needle diameter(mm)	Suture diameter(mm)	Needle–suture ratio
6/0	0.3	0.1	2.7:1
5/0	0.4	0.15	2.7:1
4/0	0.5	0.2	2.5:1
3/0	0.65	0.25	2.6:1
2/0	0.75	0.34	2.2:1
0	0.9	0.4	2.2:1
1	1.1	0.47	2.3:1
2	1.4	0.57	2.4:1

Approximation by clips and staples

Edge approximation by clips

Clips are often used instead of sutures to close skin wounds, for skin grafts and to attach prosthetic mesh in hernia repair. Clips used for this purpose are made of titanium alloys that are fully biocompatible and virtually non-reactive. The advantages of skin closure with clips include speed of application, good coaptation, minimal disturbance of the blood supply to the edges (constant occlusive force) and ease of removal. There are many clip applicators (disposable and reusable) but essentially they fall into two types: side and end applicators. The side applicators (Figure 2.10) are ideal for skin closures. They are used extensively in:

- head and neck surgery
- skin closure after peripheral vascular surgery
- skin grafting.



1 Evert skin edges.

2 Position the stapler over the centre of the everted skin edges. Close by squeezing the handles together.

3 Alternately, the skin stapler can be precocked to enable more precise placement of the staples. Simply squeeze the handles until the points of the staple are visible and the trigger is in the precocked position. The staple then can be placed exactly where desired and implanted by completing the squeeze.

4 Release the trigger to achieve staple release.

Figure 2.10 Disposable side-clip applicator for skin closure: step-by-step directions for use.

End-clip applicators deliver a clip at the end of the instrument and are ideal for the attachment of prosthetic mesh during hernia repair since they permit precise application. They are used extensively in laparoscopic hernia repair. The morbidity following use of clips for this purpose is related to the risk of nerve entrapment by the clip, especially during laparoscopic hernia repair. For this reason, in both transabdominal preperitoneal repair and extraperitoneal repair, clips are being used less frequently and only at the superior margin of the prosthetic mesh. The important factor in the prevention of recurrence is not the fixation but the size of the mesh, which must be large enough to overlap the defect by a significant margin.

Anastomosis by stapling

Anastomotic staplers are mechanical devices that enable the performance of anastomosis of hollow organs (side to side, end to end and end to side) by a single-layer technique using a treble row of evenly spaced titanium staples that join the approximated organs from the mucosal aspect while a central blade cuts out an opening inside the staple lines. In circular staplers, the blade is circular and cuts out a 'doughnut' inside the circular staple line; in linear staplers, the straight blades cut in a linear fashion the opposing walls of the two organs between the two rows of staples.

There is no doubt that staplers have been a significant technological advance since their introduction (first in Russia and Hungary) and subsequent development as fully loaded, disposable but multifire instruments in North America in the 1980s and 1990s.

The advantages of staplers, especially in surgical gastroenterology, include:

- the ability to perform safe anastomoses in surgically inaccessible or difficult areas; this is best exemplified in colorectal surgery, in which staplers have undoubtedly increased the number of sphincter-saving low anterior resections for cancer without compromising adequate distal margins
- uniform surgical results: the overall performance of low rectal and oesophageal anastomosis (effectiveness) has improved as stapling has evened out surgical skill in the execution of these difficult anastomoses
- compared with hand suturing, anastomotic stapling is considerably quicker to execute
- less risk of contamination
- avoidance of the need for occluding clamps.

The widespread use of staplers has not been an unmixed blessing. In the first instance, they have encouraged some surgeons to abandon hand-sutured anastomosis in all situations, thus being a disincentive to the acquisition and maintenance of suturing skills. Second, there are situations in which the use of staplers not only represents an unnecessary additional cost but also may enhance morbidity by an increased risk of postoperative suture-line bleeding. This is best exemplified by gastroenterostomy, where bleeding, usually from the gastric side of the staple line, is attributed to the prominent vascular supply of this organ. Finally, although stapler failure is uncommon, the integrity of the stapled anastomosis is dependent on the correct technique, which includes the correct size of stapler and staples, adequate preparation of the two anastomotic ends and anastomosis without tension. Some consider that stenosis is more common after stapled than hand-sutured anastomosis. With stainless-steel staples, corrosion and leaching of iron and other constituents of steel have been documented experimentally, although this problem has been solved with the use of titanium staples. Both types of staple produce confusing artefacts on subsequent MRI. Hand-suturing and stapling techniques are not mutually exclusive and, indeed, should be regarded as complementary since the vast majority of stapling techniques require some hand suturing for optimal results.

Underlying mechanism of tissue approximation by staples

Within the cartridge of the instrument, the staples are U-shaped. When fired, the staples grip the outer layers and assume the characteristic B-configuration that is essential for inverted coaptation of the edges without impairment of the capillary and nutrient blood flow to the anastomosis. This process works only if the surfaces of the two organs are in apposition throughout the intended anastomotic site. If extraneous tissue, e.g. fat, lies anywhere in between the stapling ring, fixation of the tissue at this point is defective and the anastomosis will leak. Thus, adequate preparation of the two ends is essential. As an additional precaution, the integrity of stapled low rectal anastomosis is tested by insufflation of air through the anus after submerging the anastomosis in isotonic saline. The selection of the correct cartridge deploying the appropriate size of staples (2.4–4.8 mm) is crucial for the execution of a safe anastomosis. The smallest staples should be restricted for the closure of blood vessels.

Types of stapling device

From the functional standpoint, staplers are either occlusive non-cutting or anastomotic cutting (linear or circular). Occlusive staplers (Figure 2.11) approximate the two walls of an organ without cutting it and are used extensively to close the stapler insertion site after a linear stapled anastomosis and in bariatric surgery.

Although circular staplers from different companies vary in detail, they are essentially of the same basic design. Circular staplers (Figure 2.12) have a detachable anvil that is inserted in one of the hollow organs and this is then closed tightly around the stem by a purse-string suture. The rest of the instrument is composed of the cartridge (containing staggered rows of staples), a docking mechanism (for the stem of the anvil), advancement mechanism for approximation of the two organs and a firing mechanism, which drives both the staples and the circular blade.

Linear staplers (Figure 2.13) consist of two docking limbs, one of which carries the cartridge containing the staplers and the other the knife mechanism. Most modern linear staplers are multifires, i.e. the cartridge can be replaced for multiple applications (up to six times). In most instances, the stapling length is fixed (e.g. 30 mm, 60 mm). Thus, if a long side-to-side anastomosis is needed (e.g. creation of an ileal pouch), sequential application of the linear stapler along the intended anastomosis is necessary. This is an entirely safe and well-validated technique. Some of the newer staplers permit adjustment of the stapling length (Figure 2.14).

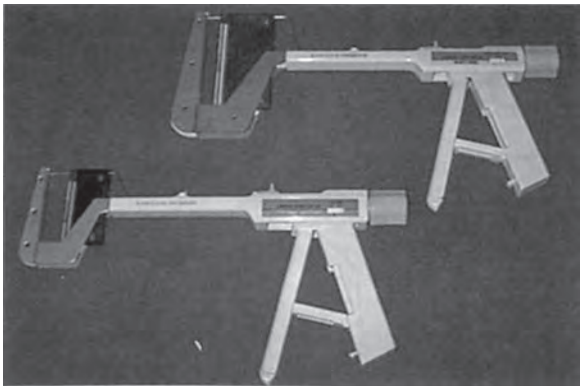


Figure 2.11 Occlusive disposable stapler: the device approximates and staples the two walls of an organ without cutting it.

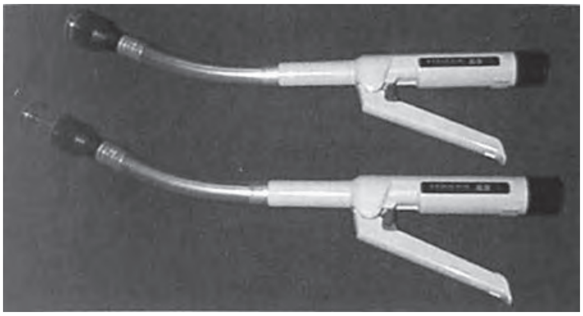


Figure 2.12 Circular cutting stapler used for oesophageal and colorectal anastomoses. The anvil is detachable.

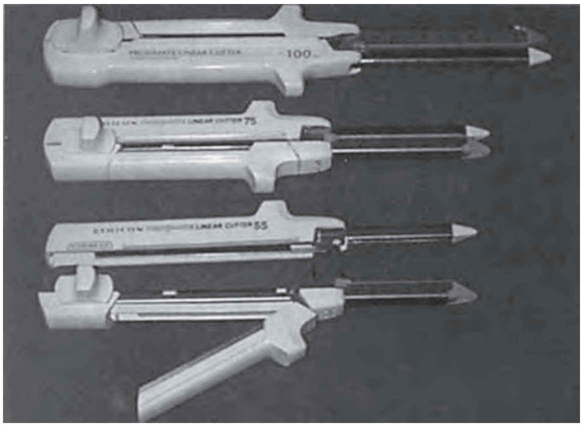


Figure 2.13 Linear cutting stapler. The instrument has two docking limbs, one of which carries the cartridge containing the staplers and the other the knife mechanism. Most modern linear staplers are multifires, i.e. the cartridge can be replaced for multiple applications (up to six times).

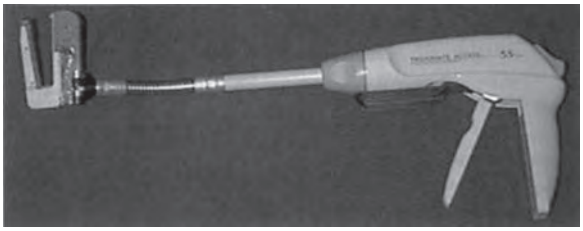


Figure 2.14 Linear occlusive stapler with adjustable stapling head.

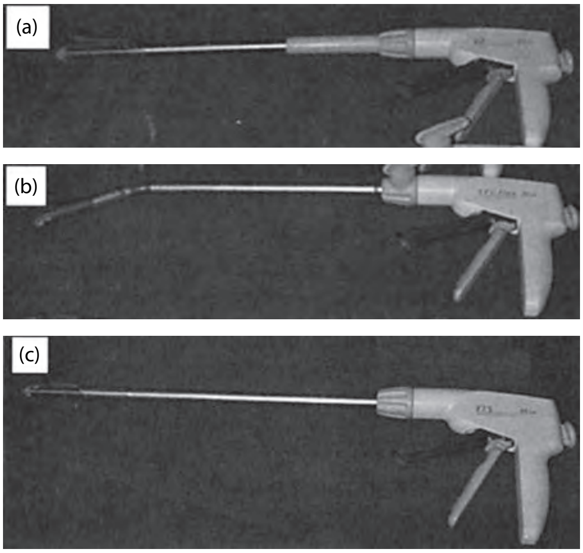


Figure 2.15 Endoscopic linear stapler: (a) occlusive stapler; (b) linear cutting anastomotic stapler with deflecting head; (c) linear cutting anastomotic stapler.

In some situations, both linear and circular staplers are used (double stapling). This is usually the case in low anterior resection in which, in the first instance, the rectum is stapled and transected by a linear cutting stapler and the stapled transected proximal colon is then anastomosed to the rectal stump by a circular stapler. This entails insertion of the anvil of the circular stapler through the antimesenteric taenia of the proximal colon and insertion of the stapler fitted with a perforating plastic trocar through the anus. The plastic trocar is made to perforate the closed rectal stump by the approximating mechanism. The plastic trocar is then removed and the two parts of the instrument are docked and approximated before the instrument is fired.

Both occlusive and anastomotic staplers have been developed for MAS (Figure 2.15). Some are powered and others have a deflecting staple head that facilitates precise application through the laparoscopic ports. Others have adjustable anastomotic lengths (Figure 2.16).

Surgical glues, adhesives and sealants

Although collectively referred to as surgical glues, a more precise nomenclature has become necessary in view of different usage/application of these biological and synthetic materials. In this context, *surgical sealants* are defined as absorbable materials used primarily to control bleeding or to seal anastomoses or skin wounds. *Surgical adhesives* denote substances with adhesive and cohesive properties stronger than sealants which are usually non-absorbable but biocompatible. *Surgical glues* have the strongest binding of all and are usually synthetic and slowly absorbable or non-absorbable. They are used for tissue approximation by binding as substitute to sutures or for mesh fixation (hernia surgery) and are exemplified by the cyanoacrylates. *Mucoadhesive polymers* stick to mucosal surfaces as they hydrate by absorbing water to form gels. Currently, they are largely used as buccal

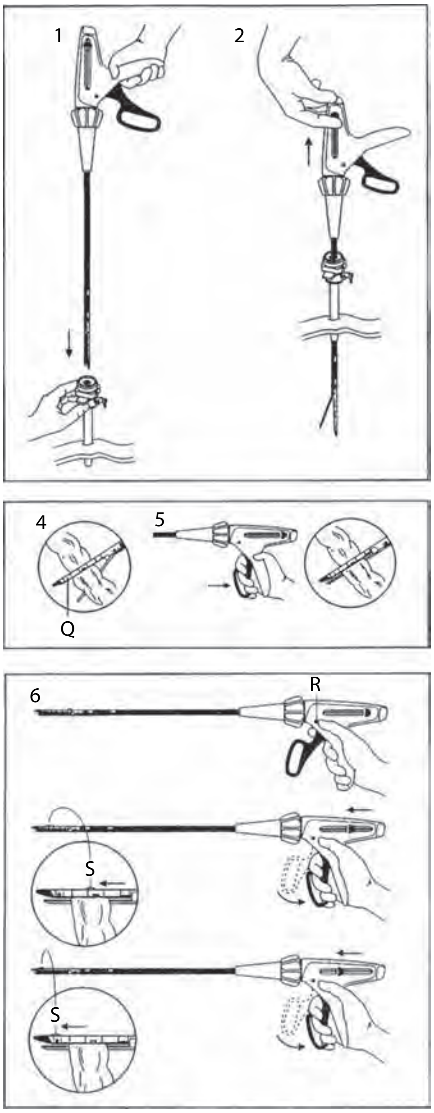


Figure 2.16 The newer generation of endoscopic cutting anastomotic stapler with adjustable anastomotic length.

mucosal drug carriers, which are assuming great importance in drug delivery. Thus a pharmaceutical drug or hormone which cannot be given orally as it is destroyed by gastric acid digestion can be incorporated in a mucoadhesive polymer that adheres firmly to the buccal mucosa, enabling absorption of the drug directly into the circulation before the polymer dissolves by hydration. There are other important emerging surgical applications involving the use of special mucoadhesive polymers, which are ferromagnetic by virtue of their iron content (as nano- or microparticles), e.g. magnetic retraction in interventional flexible endoscopy and SPLS.

The established and emerging indications for sealants, adhesives and glues indicate their widespread use and growing importance in surgical practice:

- control of bleeding during surgery (established)
- segmental solid organ resections – liver, spleen and kidneys (established)
- management of hepatic and splenic injuries (established)
- radiological embolic vessel occlusion (established)
- healing of fistulas (established)
- laparoscopic intestinal anastomosis (experimental)
- fixation of prosthetic material in hernia surgery (established)
- microanastomosis – vascular and non-vascular (established)
- repair of nerve injuries (established)
- laparoscopic vascular surgery (early clinical studies)
- skin closure (established)
- total wound closure (early clinical studies)
- endoscopic control of GI bleeding (established)
- reinforcement of suture and staple lines (established)
- sealing of pulmonary and cerebrospinal fluid (CSF) leaks (established)
- mucosal drug/hormone delivery (established)
- arthroscopic surgery – repair of cruciate and meniscus injuries (established)
- corneal surgery (established).

Fibrin-based systems

In the classic formulation, fibrin glues consist of two components – fibrinogen and thrombin plus calcium – with additives that vary with the preparation. When mixed at the point of application, the concentration of fibrinogen and thrombin influences the rate of polymerization of fibrinogen to fibrin and the strength of the final product. Some contain an antiproteinase (aprotinin), which is added with the intention of prolonging the integrity of the fibrin clot (ultimately dissolved by fibrinolysis) and increasing its adherence to the tissue. The two solutions have to be made up prior to use from the freeze-dried materials and then drawn into separate syringes that are connected to the applicator system. In essence, this consists of a Y-connection leading to a long double-lumen tube (one for each solution), with the two lumens joining at the tip where mixing of the two solutions occurs, i.e. at the point of application.

Tachocomb is a relatively new formulation of fibrin glue designed to avoid the need for making up the solutions of the procoagulant factors before use. It consists of a sponge-like membrane of equine collagen that is coated on one surface with the fibrinogen and thrombin plus calcium. The sheet is applied to the surface and then gently pressed with a saline-soaked swab until it adheres to the tissue. Tachocomb undergoes complete dissolution (phagocytosed by leucocytes and monocytes) within 6 weeks. FibFix is another formulation. It is a single-syringe formulation of fibrin glue that allows precise control of coagulation through light activation process.

The main disadvantage of all fibrin glues is the possibility of transmission of viral disease. Although all products are heat treated to destroy known viruses, there is the remote possibility of transmission of as-yet unidentified viruses. In some products, the thrombin is of bovine origin. This is a cause of some concern in view of the possibility of contamination with the bovine spongiform encephalopathy virus, and hence risk of new variant Creutzfeldt–Jakob disease.

The fibrin tissue sealants include cryoprecipitate, fibrinogen/thrombin formulations, light-activated (FibFix) and fibrin I monomer. Their advantages include elasticity and total biodegradability but they have certain disadvantages: weak adhesive polymers and a water-impermeable seal. Fibrin I monomer is now preferred to fibrinogen because it causes a minimal inflammatory response, as opposed to fibrin, requires much less thrombin for polymerization, does not require protease inhibitor and is totally biocompatible. Equally important, it produces a highly elastic fluid-repellent impermeable clot and is ready to use. Thrombin is available in two formulations: free thrombin and as thrombin nanoparticles, which are more stable, less inhibited by bacteria, pH and temperature, and have an increased storage time (6 months at room temperature).

Composite systems

The composite preparations include:

- French glue: gelatin resorcinol formalin glue
- BioGlue (CryoLife): bovine serum albumin + glutaraldehyde
- FloSeal matrix: pellets of bovine collagen coated with thrombin
- CoStasis: bovine fibrillar collagen + bovine thrombin in CaCl + fibrinogen (from patient)
- polyethylene glycol polymers: CoSeal (Baxter) and Focal Seal L and S (Genzyme)
- Tissuebond: porcine albumin + methylene blue + glycerol
- gelatin hydrogel polymer adhesives.

The polyethylene glycol polymers are totally absorbable and for internal use largely as sealants. They involve a two-stage application: primer followed by sealant which then polymerizes within 60 seconds on exposure to blue–green light (Figure 2.17). Focal Seal L is used in pulmonary resection and Focal Seal S for CSF leaks.

BioGlue is used as reinforcement of sutures/staple lines. It has been reported to cause nerve injuries and strictures in arterial surgery. Tissuemed (Tissuebond) is used as a sealant for anastomosis and consists of a dye–albumin mesh and an applicator containing porcine albumin, methylene blue (photophore to absorb light for polymerization), glycerol and water. A xenon arc lamp (Tissuemed 180 light source) is used to deliver filtered light by a fibreoptic handset activated by a footswitch. The light-activated albumin mesh then adheres strongly to tissue and grafts and is absorbable.

Cyanoacrylates

The cyanoacrylates are strong glues used extensively in flexible endoscopy to control gastric/duodenal bleeding and to fix meshes in hernia repair. They are based on 2-octyl cyanoacrylate monomer and produce heat as they polymerize (exothermic reaction), which may damage tissue and cause fibrosis. They include:

- Dermabond (Ethicon)
- Indermil (Tyco)
- Trufil (Cordis).

Their advantages include great bonding strength (tissue to tissue and tissue to prosthetic material), good vessel occlusion and excellent skin closure. They have some disadvantages, including toxicity to fibroblasts and epithelial cells, they are absorbed very slowly and thus may delay mesh incorporation, and they may carry an increased risk of infection, although this is not proven.

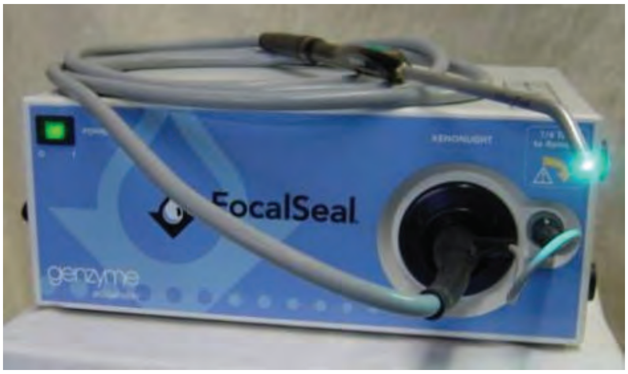


Figure 2.17 Blue–green light (xenon) for polymerization of a focal seal.

Topical materials used in surgery

These include Avitene (collagen sealant), Surgicel, Spongostan and Tacocomb. Surgicel is one of the most frequently used products in general surgery. It is a sterile, absorbable, knitted fabric prepared by the controlled oxidation of regenerated cellulose. In addition to its local haemostatic action, Surgicel is bacteriostatic *in vitro*.

Newer glues

These include biocompatible, biodegradable internal sealants based on proprietary protein engineering and consisting of a polymeric combination of synthetic silk and elastin with a cross-linking agent. They have good bonding properties, producing a flexible join. Animal studies have shown promising healing results. Photo-cross-linked hybrid ABA dendritic–linear copolyester ethers have been introduced clinically and are used in corneal surgery. CSIRO Biotechnology/University of Adelaide and spin-off company PolyNovo Biomaterials have developed a strong glue derived from the Australian frog (*Notaden*) that is biocompatible and has undergone successful clinical trials in knee joint surgery

Arrest of haemorrhage

The ability to arrest haemorrhage is an essential component of surgical skill. Haemorrhage can arise spontaneously as a result of disease or from blunt and penetrating trauma, or may be encountered during the course of an operation. The techniques for effective control of bleeding fall into three categories:

1. endoscopic: injection, thermal and photocoagulation, banding
2. radiological: embolization of bleeding vessels
3. surgical: techniques deployed during open surgery and MAS.

Control of surgical bleeding

Types of bleeding

Bleeding may be encountered during surgery or as a result of trauma (primary bleeding), within 24–48 hours after operation (reactionary bleeding) or be delayed for several days to weeks (secondary bleeding). *Reactionary bleeding* from small blood vessels within raw surfaces, minor capsular lacerations (liver, spleen) or from anastomotic suture or staple lines occurs as a result of the rise in the blood pressure after surgery. Reactionary bleeding is a special problem in laparoscopic surgery as small vessels compressed by the positive-pressure pneumoperitoneum may bleed following desufflation. Thus, careful haemostasis is essential throughout these procedures. A more serious cause of reactionary bleeding arises from a slipped ligature or clip used to secure an artery. Secondary bleeding is due to infection and is always a major problem. It is best exemplified by bleeding from infected vascular grafts, often requiring removal of the infected graft and some form of non-anatomical bypass through a non-infected region. *Internal* bleeding is defined as haemorrhage within closed body compartments or serous cavities with no visible external losses, and manifests externally as hypovolaemic shock. Occult bleeding (usually in the GI tract) refers to chronic blood loss, the exact cause of which can only be established by special investigations. It presents as iron-deficiency anaemia caused by a variety of benign and malignant disorders.

Bleeding during surgery

The primary objective is the execution of operative interventions in a bloodless field or with minimal blood loss. The advantages from this are several, the most important being reduced blood transfusion requirements, improved healing, maintenance of normal function of organ and systems, and reduction of infective complications. In many instances, bleeding during surgery is best prevented by careful dissection techniques. This is especially important in MAS, in which control of bleeding is more difficult than in open surgery and is a common indication for conversion. The following surgical methods are used for the control of haemorrhage:

- Packing: adequate packing of major or uncontrollable bleeding. During open surgery, when substantial bleeding is encountered in a deep cavity or inaccessible region, the first measure should be the application of compression by packs. This provides immediate control, allows the surgeon to obtain the necessary instruments and sutures and enables him or her to define the appropriate strategy. In addition, it gives the anaesthetist the opportunity to restore an adequate blood volume. In some situations, such as in major liver injuries, packing is the correct treatment and is preferable to hepatic resections, which carry a prohibitive mortality in this situation. Once bleeding is controlled by packs, the abdomen is closed and packs are removed or replaced 24–48 hours later. Packing is not possible in MAS, but control by compression can be achieved by grasping adjacent tissue with an atraumatic grasper, which is used to compress the bleeding area, or by pressure with a gauze swab or a tampon held by graspers. An additional port is inserted to enable application of suction and irrigation. As the assistant maintains soft-tissue/gauze compression, the surgeon approaches the bleeding area with a coagulating forceps or clip applicator in the dominant hand and suction irrigation in the assisting hand. If bleeding cannot be controlled within a few minutes, the case should be converted to the open approach, in which event soft-tissue compression is maintained while the abdomen is being opened.
- Vascular control (proximal and distal) to the operative field is a wise precaution for dissection of lesions adjacent to or adherent to major blood vessels. This entails insertion of a vascular sling around the vessels at the limits of the operative field, so that vascular clamps can be readily applied in the event of haemorrhage.
- During partial resection of solid organs (liver, spleen and kidneys), temporary occlusion of the inflow vessels is undertaken, e.g. Pringle manoeuvre during hepatic resection and management of bleeding hepatic lacerations. A careful note must be kept of the period of 'warm ischaemia' of the organ in question. The duration of warm ischaemia that the organ can tolerate should not be exceeded (45 minutes for the liver) unless the organ is cooled locally.
- Surface oozing over a wide area is best controlled by spray coagulation using the argon beam system (see Ion plasma beam).

Surgical vascular control and haemostasis

Vessels can be secured before division by a variety of techniques:

- thermal coaptation and sealing by energized systems: electrocoagulation, photocoagulation, ultrasonic sealing
- ligature and clipping
- transfixation
- vascular suturing or stapling
- application of fibrin glue.

The appropriate technique used depends on the size and the exact location. Energized coagulation by electrocoagulation, photocoagulation or ultrasonic systems (see below) is perfectly safe for vessels (arteries and veins) up to 2.0 mm in diameter but is not secure for larger vessels. Clips and simple ligatures (preferably with synthetic braided absorbable materials) are used for vessels between 2.0 and 3.0 mm in diameter. The correct application of clips, which must encompass the vessel totally and be applied at right angles to the long axis of the vessel, is crucial for safe vascular control by this technique, which is useful in deep cavity work and during laparoscopic surgery. The vessel should be divided at some distance (a few millimetres) from the clip or ligature, as the splayed-out cuff is the determining factor in preventing slippage by the force exerted by the systemic blood pressure. Overtightening of a ligature should be avoided as this will result in weakening of the vessel wall by the cheese-wire cutting effect. This is especially the case in the less pliant atherosclerotic arteries of older patients. The mechanism of closure of the lumen of an artery is a complex physical process but involves gradual reduction of the lumen (by infolding of the walls) such that blood flow ceases when the walls of the vessel collapse. When studied, the force needed to achieve this in blood vessels subjected to normal blood pressure is very small (approximately 5–7 N).

Larger arteries (3.0–4.0 mm) are more securely controlled by either transfixation or double ligature, especially if the artery arises directly from the aorta (e.g. inferior mesenteric). Vessels greater than 4.0 mm are best secured by vascular clamps on either side of the proposed section. The proximal end is then

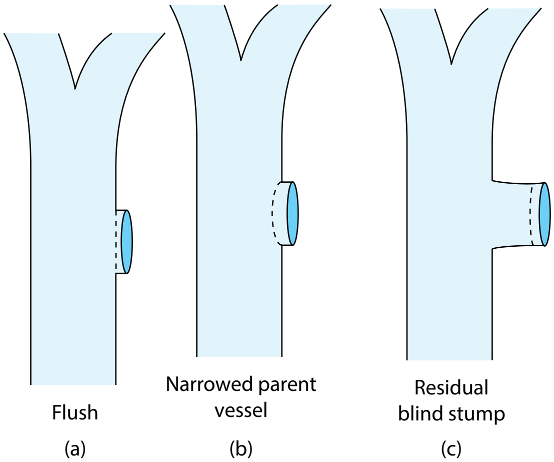


Figure 2.18 Vascular suture or a stapling of a tributary: (a) correct 'flush' control; (b) faulty because the parent vessel is constricted; (c) faulty because a blind outpouch is left. Both (b) and (c) predispose to thrombosis of the parent vessel.

sutured by 4/0 Prolene before the clamp is released. The distal end can be ligated as this will come out with the specimen. Vascular control of these large vessels by suture provides maximum security and also avoids narrowing and distortion of the parent vessel. This is especially important in securing large veins, e.g. the splenic vein at the junction with the portal vein. An alternative technique for securing large blood vessels is by vascular staplers. These are very commonly used for securing the pulmonary arteries and veins during lung resections. Whether a vascular suture or a stapling is used to secure large vessels, it is important that 'flush' control is achieved such that the parent vessel is not constricted or a blind outpouch is left. Either situation may result in thrombosis of the parent vessel (Figure 2.18).

Surgical technology

General considerations

Surgical intervention is becoming increasingly dependent on high-technology devices. This introduces new variables affecting the performance and outcome of an intervention, including malfunction of the equipment and misuse by the surgeon. The risk of errors for any particular procedure increases as the technology used becomes more complex. In addition, the technology itself is a potential source of hazard to both the staff and the patient.

Important considerations that relate to high-technology equipment used in surgical treatment include proper maintenance and training of the surgeons in safe and efficient usage. In this respect, it is essential that the surgeon understands the physics underlying the function of a particular device. Irrespective of its nature, all energized equipment that delivers energy to the patient is potentially harmful.

The power-density level reached during surgery with conventional instruments (scissors, scalpel) is a function of the sharpness of the instrument and the force applied by the surgeon. Power density is the amount of energy (mechanical, electrical, ultrasonic, photonic) applied per cm^2 of surface area of tissue and determines the effect, i.e. cutting or coagulation. In other words, the tissue effect depends on the power (W) and the surface area of contact, i.e. the effect of the same amount of energy will depend on the contact area between the instrument or probe tip and the tissue. The energy levels of the various technologies used in dissection and ablation in surgery are shown in Figure 2.19.

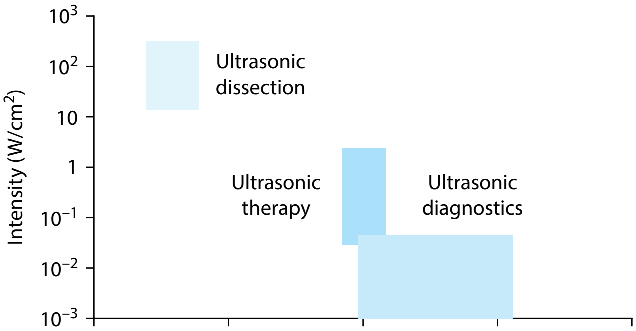


Figure 2.19 Energy levels of the various technologies used in dissection and ablation in surgery. (Reproduced from Muller and Fritzsch, *Endosc Surg Allied Technol* 1994;**2**:205–210.)

Categories of surgical technologies

Surgical technologies can be classified as follows:

- *Facilitative*: improve the efficiency of performance and reduce the level of difficulty of execution.
- *Additive*: bring technical sophistication and accuracy to surgical manipulations but are not considered essential.
- *Enabling*: make possible certain surgical procedures that would be impossible without them/open new therapeutic options.
- *Disruptive*: underpins real progress.

The designation 'disruptive technologies' was introduced by Christensen in 1997 in his book *The Innovator's Dilemma* to describe technologies that are disruptive in a positive sense because they drastically alter the way in which humans live, interact and do things. Examples of disruptive technologies in the past include the internal combustion engine, electric light bulb, railways, etc. Instances in more recent times include the internet and mobile phones. Initially disruptive technologies are more costly and less efficient than the technologies which they ultimately replace when they mature and become commonplace, such that we wonder how we managed without them. In surgery, MAS is an archetypal example of disruptive technology that, over two decades, has revolutionized surgical practice across the various surgical specialties and acted as a stimulus to radical changes in postoperative management of patients and to hospital stay, aside from being the main driver for the establishment of surgical skills laboratories. The problems with disruptive medical technologies are threefold: (1) long period between concept and realization, averaging 15–20 years, (2) maturation period and (3) training of surgeons in their safe and efficient use. Other practical issues include uptake by surgeons. The critical dynamics involved in the adoption and diffusion of new technologies are as follows:

- patients' demand for the technology (personal)
- low cost to surgeons of learning and using the procedure (professional)
- manufacturer's aggressive promotion of the technology (commercial)
- magnitude of benefit perceived by each stakeholder.

To which may be added restraint by the professional establishment and regulatory bodies such as the National Institute for Health and Clinical Excellence because of justifiable initial concerns that the new technology is both unproven and costly. This is no ready answer to this catch 22 situation as it takes time for proper assessment by appropriate clinical trials.

Image display technology

Visual perception

Human beings see the three-dimensional (3D) world with two retinas, each being two-dimensional (2D), but because of the phenomenon of retinal disparity resulting in slightly different retinal images, the brain by processing these two images enables us to enjoy true stereopsis. The human eyes perceive external electromagnetic waves of 400–700 nm (visible light) with a peak sensitivity near 555 nm. The visual apparatus sends the information from the eye to the brain after preliminary processing/coding of stimuli reaching the two retinas. The *cognitive stage* interprets the processed visual stimuli and creates an external model which enables perception. *Direct perception* relates to perception of objects in 3D space (normal everyday perception), whereas *indirect perception* is involved in viewing of pictures/images of objects. Indirect perception is thus of great importance in MAS, since the surgeon is operating from images of organs as distinct from open surgery, in which the normal direct perception is involved. The suggested mechanism involved in indirect perception is that the mind matches stimulus and memory and makes what seems to be the most likely identity and then presents this choice to consciousness. If the choice is valid this results in recognition (correct interpretation). If the choice made is incorrect, then misinterpretation results. What is important for endoscopic surgeons to realize is that the reality that is perceived is not absolute but individual, and that the correct interpretation of what is seen on the monitor by indirect perception is an acquired process which involves several factors including knowledge and experience of the displayed anatomy that essentially (irrespective of resolution and quality) remains an image and not reality.

Imaging modes

The imaging modes used in MAS are 2D, 3D and stereoscopic. 2D imaging is still the most widely used. It provides a flat image without any real depth, but, if the image quality is good (resolution, brightness and colour balance), the surgeon is able to operate safely because of pictorial (monocular) depth cues: overlapping, relative size, perspective, shadows and texture. 3D imaging is based on technologies that produce an apparently 3D image based on the after-image phenomenon (persistence of a retinal image for a few seconds using one special monitor) and the use of glasses that basically are shutting each eye alternately very quickly using polarized eyeware or liquid crystal glasses. Also there have been significant advances in 3D imaging; however, it is not true stereopsis as this requires separation of the two images and a separate telescope for each eye. In surgical practice there are only two stereoscopic systems: one used in the da Vinci robot (Intuitive Surgical) and the other in the TEM instrument (Wolfe).

A stereoscopic image contains more depth cues, enabling more accurate and efficient endoscopic manipulations. Monocular depth cues compensate somewhat for the lack of depth perception in 2D viewing and can provide comparable performance to stereoscopic (and 3D) viewing for some tasks (e.g. distance estimation). The published literature shows contradictory results on the benefits of non-stereoscopic 3D over 2D vision: some studies show better motor performances with non-stereoscopic 3D vision while others report no differences between the two imaging modalities. To some extent this controversy can be explained by the fact than all these reported comparative studies between 2D and non-stereoscopic 3D used first-generation systems, with lower resolution, brightness and inferior eye-shuttering technologies.

As the da Vinci system can toggle between stereoscopic and 2D imaging, it has enabled studies on performance by the two imaging modes. One such study demonstrated that performance using the stereoscopic imaging mode was significantly better than that achieved with 2D imaging. The time taken for the execution of all tasks was reduced by one-third and task quality improved by 25%. Dexterity was better by 25% when the surgeons used stereopsis. Accuracy based on error reduction rate improved by nearly 100%. However, other studies have reported that only the complex tasks are performed more easily and more quickly with stereoscopic viewing.

Imaging technologies: optics camera, display

The imaging chain in MAS consists of the laparoscope, the camera and the surgeon's eyes and the image display (monitor in most instances). An essential ancillary component is the light source, which is necessary for internal illumination of the body cavity and its contents.

Laparoscopes

There can be little doubt that the advent of MAS in its various forms in the early 1980s was made possible by the development of the rod–lens rigid telescope by H.H. Hopkins. Prior to then, rigid telescopes were based on the old Nitze configuration, with a series of double lenses that had to be positioned with great accuracy and the correct distance from each other in the barrel of the telescope (Figure 2.20a) with distal illumination provided by a small light bulb. These Nitze-type telescopes thus provided a dark image of poor quality without any depth of field and certainly would not have made laparoscopic surgery possible.

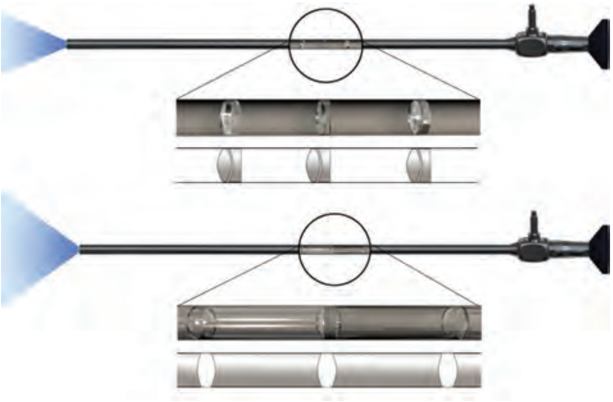


Figure 2.20 (a) Old Nitze telescope with several lens relays that had to be positioned accurately inside the barrel and had distal (bulb) illumination producing a dark often distorted image. (b) Hopkins rod–lens system that revolutionized rigid telescopes and made laparoscopic surgery possible as it produces a sharp high-quality image with depth of field.

The Hopkins system (Figure 2.20b) revolutionized endoscopy and enabled the introduction of MAS. Instead of the lens relay system, special glass rods separated by lenses placed strategically at regular intervals provide a much better aligned system capable of excellent light transmission and producing a high-resolution image with great depth of field. Moreover, the Hopkins system incorporates a light bundle that connects to a powerful light source (usually xenon) providing proximal illumination. The Hopkins rod–lens telescope is available in 10, 5 and 3 mm outer diameter. Obviously, the larger 10 mm optic produces a brighter image because of more illuminating bundles, but in practice the 5 mm telescope produces very good illumination and is used in both multiport laparoscopic surgery and SPLS. The 5 and 10 mm laparoscopes are available in three fixed viewing angles: 0°, 30° and 45°. The 0° telescope is forward viewing and is popular because it is easy to use – it sees where it points, i.e. straight ahead. Some surgeons (authors included) prefer the 30° optic because it looks down at the anatomy and, by axial rotation, either side of the median as well as up and down (normal viewing), but it does require familiarity by the camera person in its proper use. The 45° optic is rarely required in laparoscopic surgery, except perhaps in retroperitoneal operations, but it is very useful in thoracoscopic surgery.

A very recent novel laparoscope known as the Endocameleon (Storz) is capable of altering the viewing angle and is thus very useful (Figure 2.21); to date, it is only available as a 10 mm telescope.

Optoelectronic telescopes

These are different from the standard rigid laparoscopes (which couple to the camera) because the light sensor is inside the telescope, behind the lens, and indeed the telescope incorporates all other functions: light transmission through fibreoptic bundles, control of distal end, focus and zoom. All the modern flexible endoscopes belong to this category and produce high-quality imaging. They are used in NOTES and in SPLS, in which a flexible optic reduces internal collision between the telescope and the instruments. To date, they have not proved popular among surgeons in multiport laparoscopic surgery.

Cameras

The modern laparoscopic cameras are based on chip sensor devices of two basic types: charged-couple device (CCD) or complementary metal oxide semiconductor (CMOS).

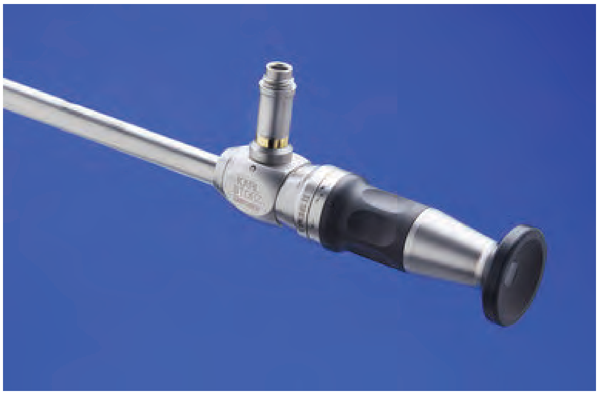


Figure 2.21 Endocameleon (Storz).

Boyle and Smith developed the CCD, for which they were awarded the Nobel Prize for Physics in 2009. Initially, the CCD was intended as a shift register (type of memory device). Tompsett *et al.* subsequently demonstrated that it could also accumulate charge via the photoelectric effect, thereby creating images. The CCD ushered in a new era of miniature high-resolution cameras that replaced all the existing imaging systems. CCD and, more recently, CMOS chips are used either at the distal end of an endoscope, forming digital (optoelectronic) endoscopes, or in a lightweight camera coupled to a Hopkins rod–lens telescope with output via a camera control unit to an external image display. Both CCD and CMOS chips are pixelated metal oxide semiconductors with an epitaxial layer of silicon as the photodetector. They accumulate signal charge in each pixel that is proportional to the illumination intensity. When exposure is complete, the CCD transfers each pixel's charge packet sequentially, as a shift register to a common output structure, which converts the charge to a voltage and sends it to the display. In a CMOS sensor, the charge-to-voltage conversion takes place in each active pixel sensor. CMOS chips have the advantages of high integration and functionality, small size and low power consumption – features which are particularly suited for small systems, mini-endoscopes, mini-robots and wireless capsule endoscopy (PillCam) introduced by Given Imaging in the 1990s.

The images from the CCD or CMOS sensors are black and white, and colour imaging is produced by filters. The first-generation CCD cameras had one chip overlaid by a Bayer

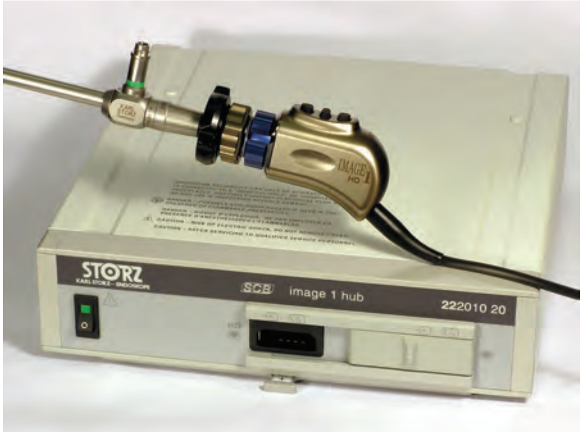


Figure 2.22 Modern progressive high-definition television camera producing progressive imaging at 1080 lines (Storz).

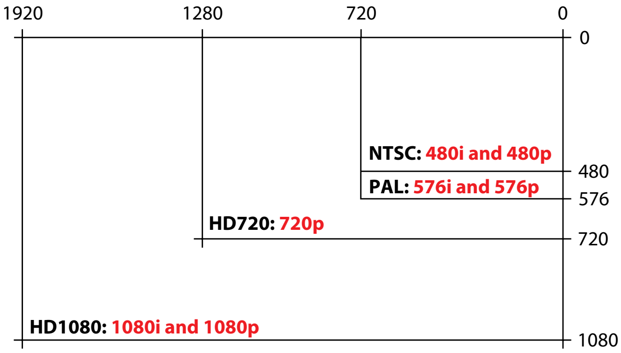


Figure 2.23 The higher resolution range of high-definition (HD) systems (camera and monitors). The HD systems range from 720 to 1080 lines. NTSC, American system; PAL, European system.

mask, which provided each square of four pixels with three filters: red, blue and two green (RGB). The second generation of laparoscopic cameras used nowadays have much better colour separation and luminance because they are equipped with three CCDs and a dichroic beam prism that splits the image into its red, green and blue components with each of the three CCDs being configured to respond to one colour. The three-chip camera has superior colour rendering and higher light sensitivity for a given aperture size. The best cameras are now equipped with three high-definition (HD) CCD chips and produce the best image quality (Figure 2.22).

Image display technology

Monitors

Standard

Cathode ray tube (CRT) monitors are being replaced by high-resolution back-lit liquid crystal display (LCD) monitors, and increasingly HD television (HDTV) systems are preferred because of their higher resolution. Although the aspect ratio (length to height) of HDTV monitors is usually 16:9, compared with 4:3 or 5:4 with standard definition (SD) monitors, there are low-end HDTV systems (720 lines) with aspect ratios in the SD range.

High definition

An HD image can contain up to five times the picture information contained in SD monitors. SD monitors consist of approximately 500–600 horizontal lines of pixels depending on the local transmission system used, whereas HD transmissions use at least 720 to over 1000 lines (Figure 2.23).

There are two HD systems: interlaced, denoted by the suffix ‘i’, and progressive, denoted by the suffix ‘p’. Each frame of the image display is made up of lines (in turn made up from pixels) that are built from above downwards on the screen and make up the picture, with the frames changing very rapidly (frame refresh rate is usually 30 Hz). In the interlaced system, the odd-numbered lines forming the picture are displayed first from above downward, e.g. 1–3–5–7 etc., until the bottom of the screen is reached, when the system starts again from the top and fills in the gaps with the even-numbered lines. The progressive system instead projects all the lines forming the screen picture simultaneously. The progressive system provides better quality imaging when the scene is changing rapidly and is sometimes referred to as full HD.

Organic light-emitting diode technology and displays

An organic light-emitting diode (OLED) is a light-emitting diode (LED) in which the emissive electroluminescent layer is a film of organic compounds/material that emit light in response to an electric current, the layer of organic semiconductor material being connected between two electrodes. There are two types of OLEDs: (1) those based on small molecules (quantum dots) and (2) those employing polymers, the latter being preferred for TV monitors. OLED displays can use either passive-matrix (PMOLED) or active-matrix (AMOLED) operating schemes. AMOLEDs require a thin-film transistor backplane to switch each individual pixel on or off, but enable higher resolution and larger display sizes.

Organic light-emitting diode and liquid crystal display systems

OLED displays do not need a backlight. In low ambient light conditions, such as in MAS operating rooms, OLED screens can achieve a higher contrast ratio than an LCD irrespective of whether this uses CRT or back-lit LED monitors. OLED displays have more control over colour rendition because they produce only pure colours when an electric current stimulates the relevant pixels. The OLED primary colour matrix is arranged in red, green and blue pixels, which are mounted directly onto a printed circuit board with each individual OLED element being housed in a special ‘micro-cavity’ designed to reduce interference from ambient light, thereby improving colour contrast. The thickness of the organic light-emitting layer is adjusted to maximize light for each of the primary colours (red, green and blue).

OLEDS require a low drive voltage and thus their power consumption is low. An LCD pixel works by applying different voltages to three electrodes, which each align to a small area of liquid crystal material to allow a certain amount of backlight to shine through a red, green and blue colour filter. These three subpixel colours then combine to form the final light output for one pixel, with the liquid crystal material acting as a variable shutter for each of these colour components. An OLED pixel works differently: light is generated directly in each subpixel when different currents are applied to the three diodes, each diode then emitting red, green or blue light. The final OLED light output results from the combination of the subpixel colours, similar to LCD.

Potential of organic light-emitting diode displays for minimal access surgery

OLED monitors provide superior quality for MAS than CRT and LCD systems for the following reasons:

- Energy efficient: use less power as they have fewer layers and no light is wasted by shining against non-transmitting liquid crystal material. Overall, OLED displays have average energy savings over LCD systems between 0% and 40% depending on use and type of LCD.
- Thinner lighter displays: as they dispense with liquid crystal, colour filter and associated layers required for LCD. The average thickness of OLED displays is 2.6 mm.
- Enhanced colour rendition: OLEDs give off a truer and wider range of colour than the LCD displays, which shine a backlight through a colour filter. OLED displays have much higher contrast ratios as a result of much richer blacks.
- OLED displays are flexible and can be transparent, and do not require glass layers to hold the liquid crystal intact.
- The colour quality of the image is the same regardless of viewing angle, as distinct from LCD systems in which the colour mix will change at wider viewing angles – they can be viewed from different angles.
- In low ambient light conditions, such as in MAS operating rooms, OLED screens can achieve a higher contrast ratio than LCD displays.
- Higher refresh rates than LCD displays, and thus more suitable for 3D viewing. The refresh rates (number of times per second a display redraws data) required for some 3D systems is twice that of an LCD to achieve the same image quality. The OLED pixels have a faster response rate as they are directly driven, whereas the LCD lighting element passes through the liquid crystal shutter.

Most OLED displays are currently used for entertainment and in higher end mobile devices, but Sony has just produced an OLED monitor for MAS that is undergoing clinical evaluation. It is likely that OLED displays will eventually replace LED and CRT systems for MAS and are likely to enable the development of image projection systems, dispensing with the need for ceiling-mounted or free-standing monitors.

Projection systems

These aim at projecting the image of the operating field on top of the patient in front of the patient, thus enabling 'gaze-down imaging' such that the surgeon can see his or her hands. Experimental studies have demonstrated that gaze-down frontal imaging on top of the patient improves both task efficiency (less time to complete the surgical task) and task quality (better execution). The first projection imaging system for MAS (ViewSite) failed and is no longer in production because of the poor quality of the projected image and indifferent ergonomic design. Since then, there have been significant developments in projector technologies and considerable research and development on disposable sterile screen material for image projection. Such a prototype system being developed in Dundee, UK, is shown in Figure 2.24.

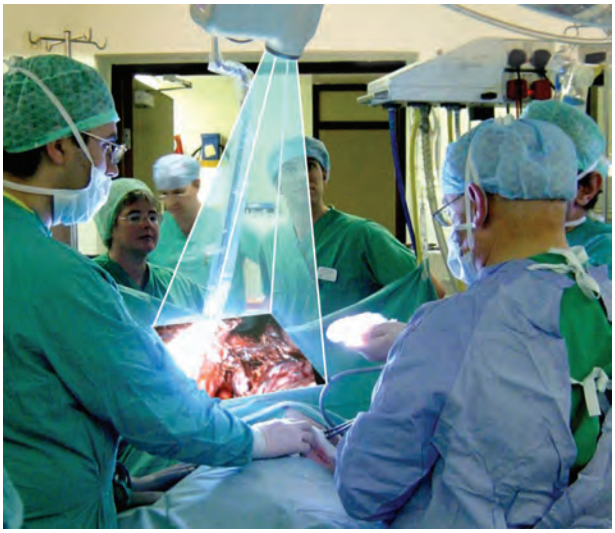


Figure 2.24 Projection image display for minimal access surgery. The image is displayed on a sterile screen on top of the patient, enabling gaze-down imaging.

It is likely that recent advances, including plastic OLED displays (2–3 mm) and laser holography capable of 'suspending' the image on top of the patient, will accelerate the development of image projection for MAS.

Ergonomics of imaging

Aside from the quality of the image display, there are important considerations that pertain to the ergonomics of viewing by the surgeon that are often overlooked but that are important because they affect both the quality and efficiency of the execution of the operation. These ergonomic principles have three components: (1) viewing position of the surgeon, (2) eye gaze and stance of the surgeon and (3) distance from the monitor.

Viewing position of the surgeon

As stressed at the start of this section, the surgeon is an integral, and indeed the most important, component of the imaging chain of MAS. Surgeons must be positioned such that they face the monitor directly and do not have to turn their head to either side for viewing the image on the screen. They must have clear vision immediately ahead with no obstacles in the viewing path.

Eye gaze and stance

Humans perform precision tasks best when they look down at the task, and this applies to all professions with MAS being no exception. Experimental studies have confirmed that the best execution of surgical endoscopic tasks is achieved when the surgeon looks down at the image (gaze-down viewing), and performance deteriorates when the surgeon operates looking up, or to the left, or to the right, to see the image of the operating field on the screen. With the current monitor-based image display systems, gaze-down viewing is problematic, especially for short individuals, but these surgeons are advised to stand on a platform. Optimal gaze-down viewing can only be achieved by projection imaging on the patient (see above).

Until such systems come on stream, the best arrangement is by an LCD display adjustable to the individual height of surgeon so that he or she can look down at it.

Distance of surgeon from monitor

This of course varies with the size of the monitor (conventionally measured as the diagonal of the image display in inches), but there is in addition great variation between surgeons. In studies with a 21 inch SD monitor, the maximal viewing distance among trained laparoscopic surgeons with corrected vision ranged from 139 to 303 cm, the minimum viewing distance (how close the surgeon is to the monitor and still see the image clearly) ranged from 90 to 183 cm, and the individual monitor optimal view distance (the distance the surgeon preferred for optimal viewing) ranged from 90 to 303 cm. This study demonstrates that each surgeon has his or her optimal viewing distance, with some preferring close-up viewing and others not.

Energized surgical dissection technology

High-frequency electrosurgery

The underlying principle involves the passage of an electric current through the tissue by means of a potential difference (voltage). The resultant flow of electrons excites the tissue molecules, notably water, creating heat energy, which causes water evaporation and tissue coagulation. High-frequency (HF) currents (greater than 100 000 Hz) are used to minimize the risk of electrical shocks. The characteristic of the applied current can be varied from sine wave to modulated, to favour cutting or coagulation, respectively (Figure 2.25). HF electrosurgery can be monopolar, bipolar or quasibipolar.

Monopolar high-frequency electrosurgery

Here, the current escapes from the electrode tip, dissipates into the receptive tissue and exits through the grounding pad (patient neutral electrode; Figure 2.26). The neutral grounding pad should have a large, even surface and is applied to the patient's upper thigh, covering a large area of clean, hairless skin to ensure a low current density. Special efforts should be made to ensure that the patient's skin does not come in contact with metal components of the operating table, chucks and restraining devices, and damp clothing. The connecting cables to the neutral and active electrode (appliance used by the surgeon) must be in perfect condition with respect to their insulation and connector plugs.

The potential hazards of HF electrosurgery include unwanted discharge and absorption of low-frequency currents (shocks), HF current burns (to patient and staff), sparking to unwanted areas (collateral damage) and capacitive coupling. Electrical current leakage can also cause malfunction of pacemakers and monitors. The monopolar HF unit must not be operated in the vicinity of an electrocardiogram electrode (minimum distance of 15 cm). Modern electrosurgical generators are smart devices in that they incorporate a microprocessor with sensor electronics that provide the necessary feedback from the electrode–tissue interface to the computer inside the generator. Used correctly, these smart generators adjust the power output to the lowest level necessary to achieve the effect. In addition, the strict limitation to the purposeful voltage range (200–500 Vp) preserves the integrity of the electrodes and increases the safety margin.

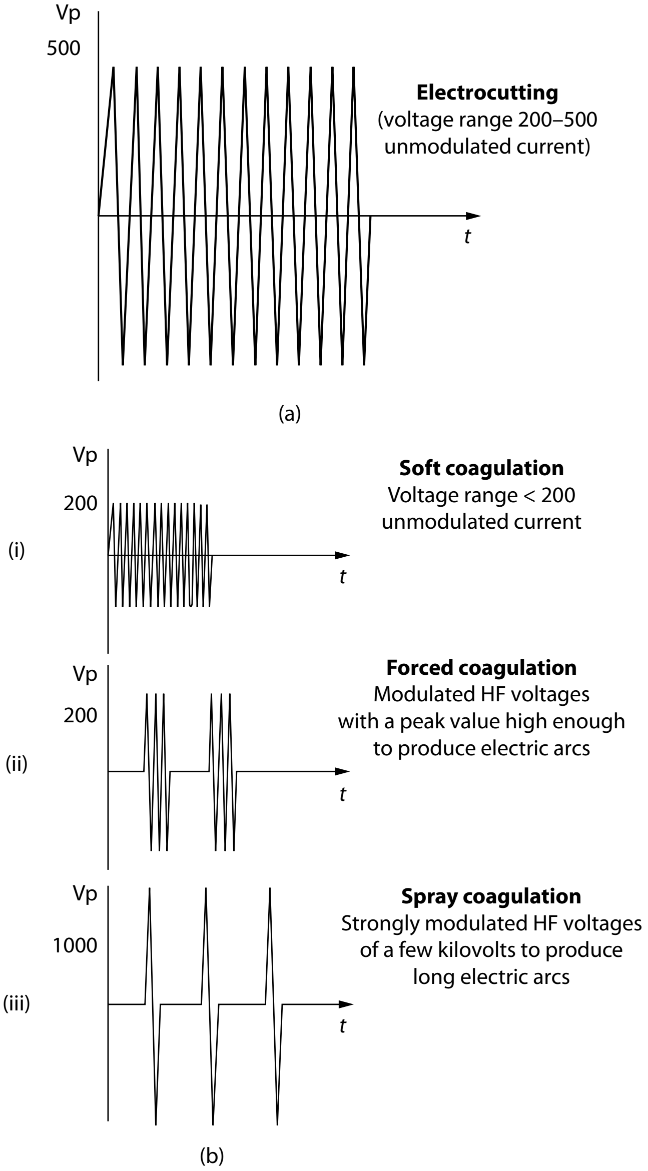


Figure 2.25 Characteristics of the applied currents in high-frequency (HF) electrosurgery. (a) Unmodulated continuous sine wave in the voltage range 200–500 V for electrocutting. (b) (i) Unmodulated current with a voltage output below 200 V for soft coagulation; this voltage does not generate electric arcs. (ii) Modulated HF voltages with a peak value high enough to produce electric arcs for forced coagulation. (iii) Strongly modulated HF voltages of a few kilovolts to produce long electric arcs for spray coagulation.

Electrocutting

HF monopolar electrocutting is achieved by the use of a continuous sine wave (unmodulated) current (Figure 2.27) of sufficient voltage (200–500 V) to produce electric arcs at the tip of the active electrode. These arcs cause an immediate vaporization of tissue in contact with the electrode, creating a series of 'cellular mini-explosions' that collectively result in cleavage. If the voltage drops below 200 V, tissue cutting cannot be achieved since electric arcs are not generated. Conversely, if the voltage rises above 500 V, the electric arcs produced are so intense that

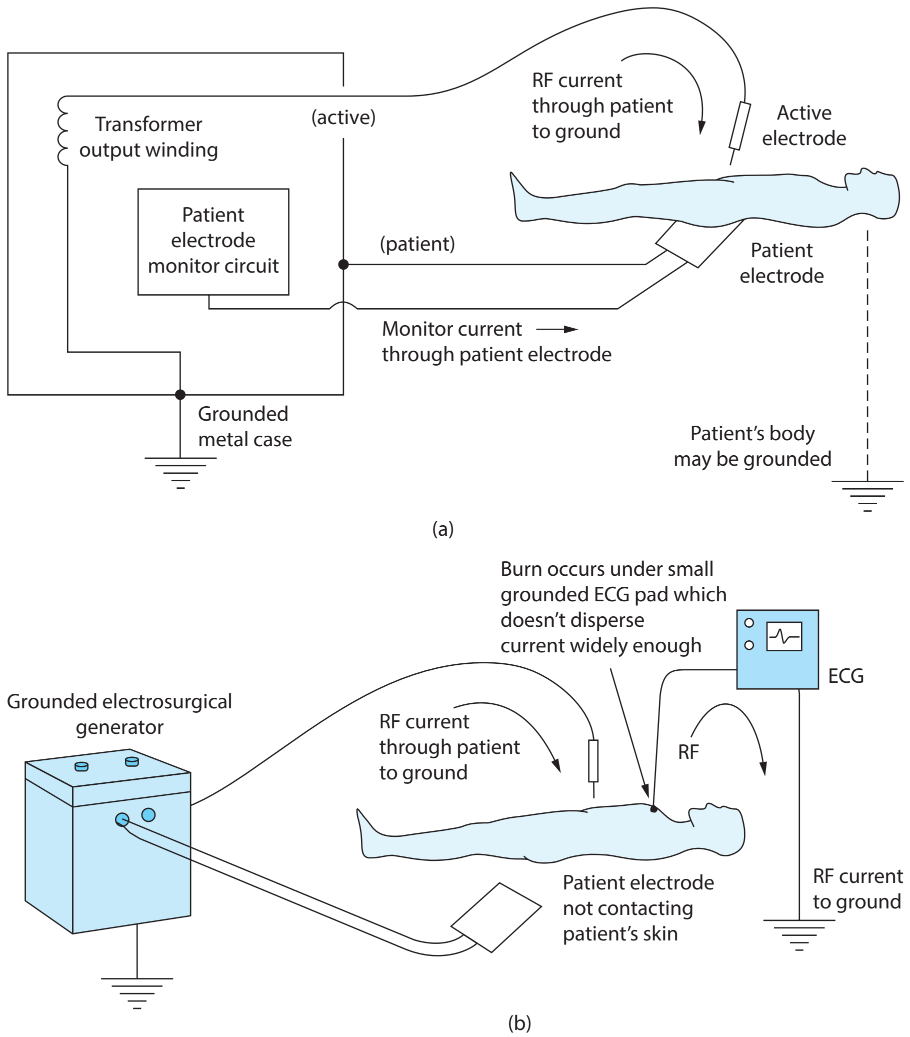

Transformer output winding
(active)
RF current through patient to ground
Active electrode
Patient electrode monitor circuit
(patient)
Patient electrode
Monitor current through patient electrode
Grounded metal case
Patient's body may be grounded
(a)
Burn occurs under small grounded ECG pad which doesn't disperse current widely enough
ECG
Grounded electrosurgical generator
RF current through patient to ground
RF
Patient electrode not contacting patient's skin
RF current to ground
(b)

Figure 2.26 Sequence of events leading to a thermoelectric burn of the patient following faulty application of the indifferent electrode of a diathermy generator using the conventional earthing system. (a) Correct application of the indifferent (patient's) electrode; (b) faulty application of the patient's electrode.

the tissue is carbonized and the electrode may be damaged. Within the safe operating range, the depth of coagulation of the cut edges increases with increasing voltage and intensity of the electric arcs. In practice, the depth of coagulation during electrocutting is determined by the setting of the HF output power and the degree of modulation of the current (blend or mix). It is also influenced by three other factors:

1 thickness of the cutting electrode
2 rate and depth of cutting
3 impedance of the generator.

With a conventional HF unit having an impedance of 250 W, fluctuations in the current or output voltage and intensity of the electric arcs are produced by variations in the depth and rate of cutting. This may lead to carbonization of the tissue along the cut edges, particularly at the beginning and end of the cut. This problem is obviated by smart HF generators that incorporate microprocessor-controlled automatic circuits with sensory electronics referred to previously.

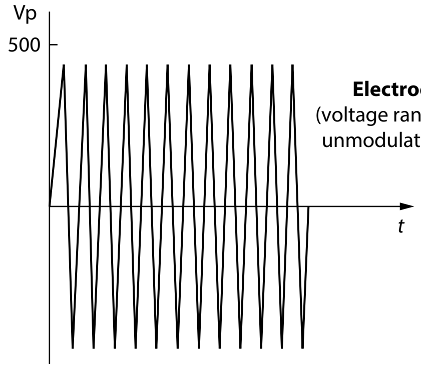


Figure 2.27 Characteristics of the applied currents in high-frequency electrosurgery. Unmodulated continuous sine wave in the voltage range 200–500 V for electrocutting.

Electrocoagulation
When an HF-modulated alternating electric current is applied specifically for heating living tissue, the temperature rise is proportional to the specific electrical resistance of the tissue, the duration of current flow and the square of the root-mean-square of the electrical current density. Largely because of the irregular current density distribution in the tissue, the temperature rises at different rates in various zones within the tissue. As the current density is largest in the zone of contact between the tissue and the electrode, the maximum temperature is reached

there, with the temperature decreasing proportionately with the distance from the contact area. The safe deployment of monopolar electrocoagulation depends on this phenomenon, since the further away the tissue from the contact area, the less likely the damage.

When the temperature of the tissue near the contact surface reaches boiling, a layer of vapour forms between the tissue and electrode, impeding current flow. Thereafter, the sequence of changes depends on the peak voltage. If this is less than 200 V, the coagulation process slows down until the tissue next to the electrode has dried out, when current flow ceases. If the current is not switched off, the coagulum becomes adherent to the electrode. When this happens, removal of the electrode will dislodge the coagulum and precipitate renewed bleeding. If the peak voltage exceeds 200 V, once the tissue next to the contact area has dried out, electric arcs are produced that carbonize and puncture the coagulum, thereby causing the coagulation process to continue unabated until the generator is switched off or the dried-up coagulum becomes so thick that it resists further puncture by the electric arcs.

There are three coagulation modes that can be used:

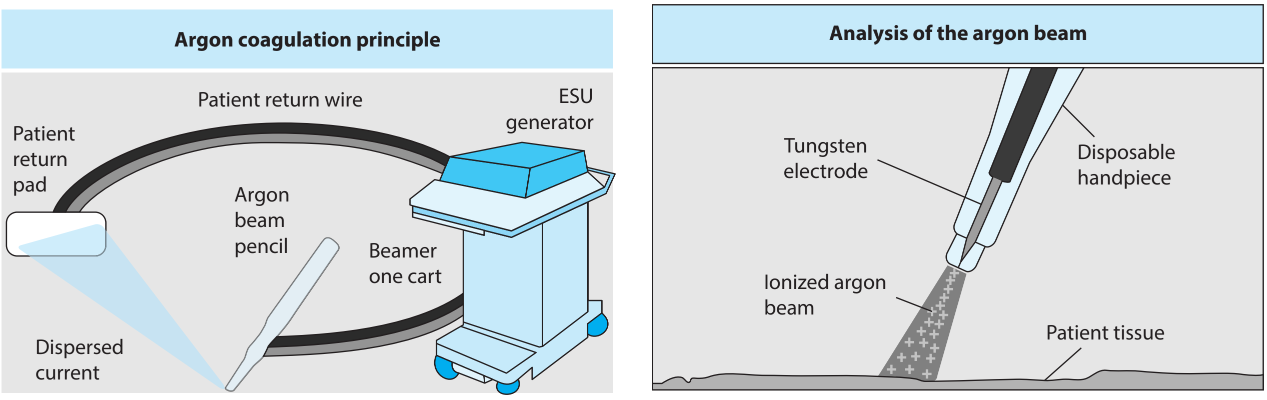
1 soft
2 forced
3 spray.

Soft coagulation is the safest for both open and laparoscopic surgery. As the peak voltage is less than 200 V, no electric arcs are generated between the electrode and the tissue. It results in desiccation of the tissue with shrinkage and no charring. *Forced coagulation* deploys high peak voltages (>500 V) to generate electric arcs to achieve deep coagulation. This may be needed in vascular areas but the risk of collateral damage is substantially higher. Forced coagulation should never be used in close proximity to important structures such as the bile duct or ureter. *Spray coagulation* is a non-contact mode in which long electric arcs are intentionally generated by strongly modulated HF voltages (a few kilovolts) to surface coagulate raw and bleeding areas or to achieve haemostasis from inaccessible vessels. Spray coagulation is used in urology during transurethral resection of the prostate.

Ion plasma beam coagulation

This is a modification of monopolar HF electrocoagulation. It utilizes a plasma of ions of an inert gas (most commonly argon) to deliver the electrical current and is directed as a focused beam to the desired target. The argon gas beamer works with certain specific generators, which are capable of providing a spray coagulating current. The electrical arcs from this current produce a conductive channel of argon ions (the plasma) in the centre of the gas cone. It thus provides non-contact coagulation with minimal charring and depth of tissue necrosis under the coagulum. The argon ion plasma appears as a blue lightning flame that darts to other wet areas of the tissue surface once the incident zone has been desiccated, hence the limited depth of coagulation. Other advantages of the system include the displacement of blood by the gas spray, thereby permitting a more direct application of current to the bleeding area, and cooling around the coagulating zone by the outer layers of the gas, thereby minimizing lateral thermal damage. There is an optional setting of the gas flow in litres in relation to the generator power output, i.e. the greater the power, the greater the gas flow required. When operating on vascular organs such as the liver, spleen and kidneys, the argon flow rate must not exceed 10.0 L/min because of the risk of gas embolism.

The delivery hand piece (disposable or reusable, flexible or rigid) essentially consists of a tube containing the energizing copper wire or tungsten wire electrode (Figure 2.28). Ion plasma coagulation can be used in open surgery, MAS and interventional flexible endoscopy to control bleeding from surface lesions, and is especially useful in arresting haemorrhage from large oozing surfaces, during liver resections and partial splenectomy or nephrectomy. During laparoscopic and thoracoscopic surgery, one of the ports must be kept open during deployment of ion plasma coagulation, to avoid a dangerous rise in the intra-abdominal and intrathoracic pressures, respectively.



(a) (b)

Figure 2.28 (a) The argon gas beamer works with certain specific monopolar generators, which are capable of providing a spray coagulating current. A ground pad is essential for the return current and the patient must be insulated from any metal component or wet surface. (b) The electric arcs from this current produce a conductive channel of ionized argon gas in the centre of the gas cone, resulting in non-contact precise coagulation with minimal charring and depth of tissue necrosis underneath the coagulum.

Bipolar electrocoagulation

Some dedicated bipolar generators operate only with bipolar probes, whereas others are capable of operating in both the monopolar and bipolar modes. In bipolar electrocoagulation, the current crosses between the two prongs of the electrode and returns to the generator without flowing through the patient. Thus, a patient ground (neutral) pad is not required and bipolar electrocoagulation is undoubtedly safer than the monopolar variety. Heating of the tissue in bipolar electrocoagulation is confined to the area between the two ends of the probe and thus collateral thermal damage is not possible. When using bipolar electrocoagulation, the minimum power necessary is used and every precaution must be taken to avoid contact between the active electrode and metal instruments. The tips of the bipolar electrode must be kept clean, as fouling with encrusted coagulum impairs the coagulation efficiency.

Although bipolar generators do not require a neutral patient ground pad, combined machines, which incorporate monopolar and bipolar circuits, do, and the safeguards described previously with regard to the safe application of the neutral ground pad must be strictly adhered to.

Quasibipolar electrocutting

This is designed to achieve electrocutting without the unmodulated current passing through the patient. Although standard monopolar/bipolar generators are used, specially designed cutting probes, which incorporate active (cutting) and return elements, are employed to apply the unmodulated current (Figure 2.29). The system is not, strictly speaking, bipolar and some current leakage to the patient is possible, especially if the return component of the electrode is fouled by charred tissue debris. A neutral ground pad is thus advisable.

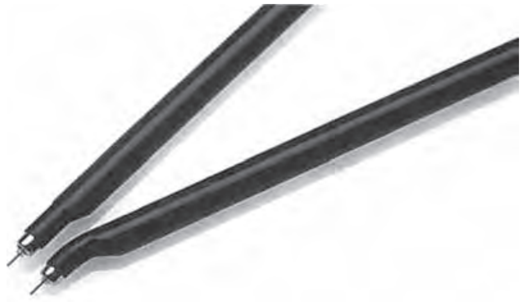


Figure 2.29 Bipolar electrocutting knife. The cutting needle can be retracted inside the instrument and the 'return' electrode.

Chemical composition of electrosurgical smoke

Electrosurgical smoke contains charred debris, cell-sized fragments, breathable aerosols and various complex organic molecules. Electrosurgical smoke used to be considered sterile but this has been questioned, particularly in relation to viral agents. The same applies to the smoke (plume) caused by lasers. At least 21 toxic or carcinogenic compounds have been positively identified in electrosurgical smoke (Table 2.4).

The use of smoke-evacuation devices and safety precautions against inhalation by theatre staff has been recommended. In endoscopic surgery, compounds produced by the high-temperature pyrolysis of protein and fat are absorbed into the systemic circulation of the patient and lead to an increased level of carboxyhaemoglobin and methaemoglobin in patients after prolonged laparoscopic operations. *In vitro* studies have shown that high concentrations of electrosurgical smoke are cytotoxic and low (sublethal) concentrations stimulate white cells and the vascular endothelium.

Ultrasonic dissection

Ultrasound waves are mechanical waves that are not audible by the human ear, i.e. have a frequency above 20 000 cycles per second (20 kHz) and propagate in matter but not in air (hence the need for acoustic coupling by conductive gel during contact diagnostic ultrasound). Ultrasonic dissectors impart a higher power density than mechanical non-energized dissection but utilize less energy than HF electrosurgery or laser surgery. As a result of this, they cause less heating of tissue with reduced penetration. The ultrasonic frequency range used for surgical dissection ranges between 20 and 60 kHz. The tissue effects (separation, cutting, coagulation) are due to the high power density (around 100 W/s) compared with diagnostic ultrasound, which, because of the low power density (0.01 W/s), has no discernible effect.

Basic design of ultrasonic dissectors

All ultrasonic dissectors, irrespective of make, consist of (1) an electrical generator, (2) a piezoelectric transducer and (3) the dissection instrument (disposable or non-disposable). The transducer and dissection instrument are incorporated in the hand piece, which is connected to the generator by an electrical cable. The generator supplies the electrical energy to the system.

Piezoelectric transducers

These are ferroelectric ceramic crystals that vibrate (expand and contract rapidly), thereby producing ultrasound waves

Table 2.4 Chemicals identified in electrosurgical smoke

Hydrocarbons	Nitriles	Amines	Aldehydes	Miscellaneous
2,3-Dihydroindene	3-Butenenitrile	Pyrrole	3-Methyl propanol	2-Methyl furan
1-Decene	Benzonitrile	6-Methyl indole	3-Methyl butenal	5-Dimethyl furan
1-Undecene	2-Propylene nitrile	Indole	Furfural	Hexadecanoic acid
Ethynyl benzene			Benzaldehyde	4-Methyl phenol
Ethyl benzene				
Toluene				

when energized by an electric current. The vibration frequency (Hz) varies with the extent of polarization of the crystals. By contrast, when piezoceramic crystals are compressed, they generate an electrical current, i.e. convert mechanical into electrical energy. It is this latter effect that is used in many household and car appliances, e.g. ignition systems. The former effect, conversion of electrical to mechanical (ultrasound) energy, is used in both diagnostic ultrasound and ultrasonic dissection.

The vibrations generated by the piezoelectric transducer are conducted by a metal rod, the length, diameter and shape of which influence the conducting efficiency to the tip of the instrument. A number of silicon rings fixed at specific vibration nodal points (Figure 2.30) prevent the rod from touching the outer protective sheath. Alternatively, the rod is covered in a Teflon sheath. An efficient system transmits most of the vibration energy to the tip of the instrument (hook knife, shears, ball, etc.) with little heating of the shaft of the instrument.

Low-power ultrasonic dissectors

These operate at a frequency of around 25 kHz and the hand piece, consisting of the transducer and free (unsheathed) rod, is non-disposable. The system allows irrigation and suction during activation, and hand pieces for both open and laparoscopic surgery are available (Figures 2.31 and 2.32). Low-power ultrasonic dissectors are the safest form of energized dissection. They do not coagulate or cut, but only cleave cells with a high water content by a process called cavitation: the intracellular water, when subjected to the pressure waves, vaporizes, forming vacuoles, which then resonate with the vibrating rod, leading to implosion of the cell. These low-power ultrasonic dissectors cannot cut fibrous organized structures (arteries, veins, ureters, etc.) and impart very little, if any, collateral damage. They are used extensively for hepatic resections, where they cleave the hepatic parenchyma, exposing the intrahepatic ducts and vessels for clipping or ligation.

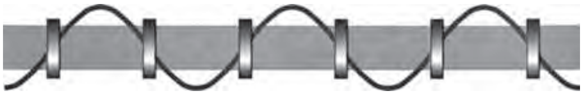


Figure 2.30 A number of silicon rings fixed at specific nodal points prevent the vibrating rod from touching the outer protective sheath of the hand piece of the ultrasonic dissector.

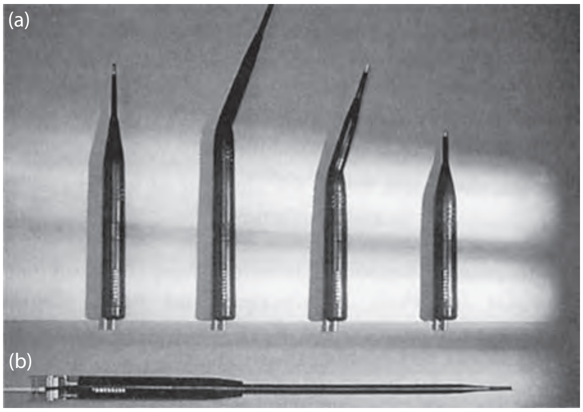


Figure 2.31 Selector probes for low-power ultrasonic dissection: (a) for open surgery; (b) for laparoscopic surgery.

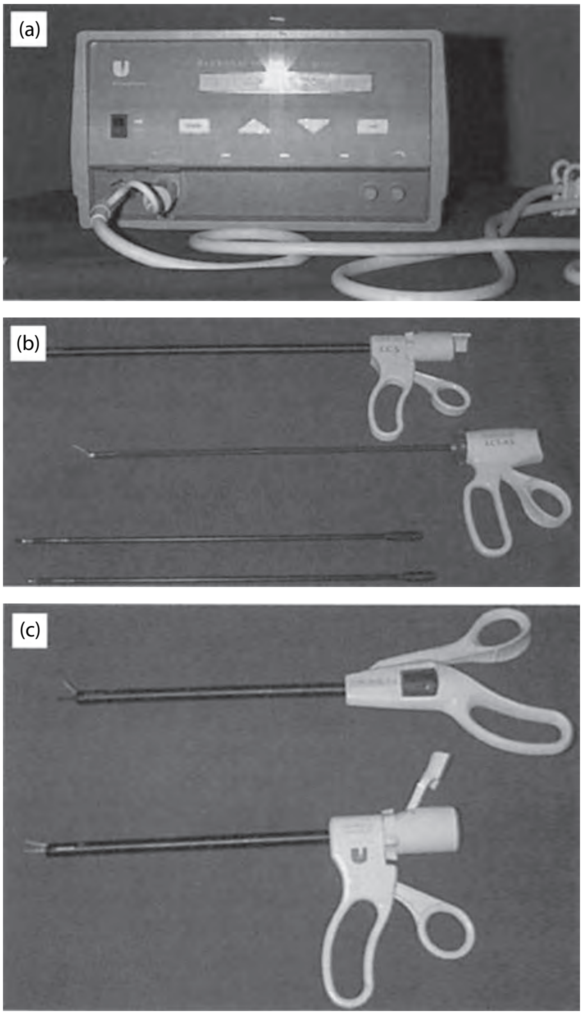


Figure 2.32 High-power ultrasonic dissection system (Ethicon). (a) Generator; (b) transducer and hand pieces for endoscopic surgery; (c) hand pieces for open surgery.

High-power ultrasonic dissectors

High-power ultrasonic dissectors have become very popular for both laparoscopic and open surgical interventions in view of their efficiency in effective coagulation and cutting, thus reducing the instrument traffic and the duration of the procedure. Instead of smoke, they produce mist that can obscure the view in laparoscopic work, especially when the sprayback fouls the lens of the telescope.

These devices operate at a frequency of 55.5 kHz and thus deliver a very high vibrational energy to the tissue such that significant heating (as much as 120°C at the instrument tip), deformation and friction effects occur at the instrument–tissue interface. Cutting is mainly due to 'tissue sawing' (high-speed frictional deformation) coupled with linear compression of the tissue (by the surgeon). Although very efficient, these high-power ultrasonic dissectors can cause collateral tissue damage, especially when the generator is used at maximum power. Animal experiments have demonstrated significant damage (on subsequent histology) to the walls of the aorta, bile duct, etc., after high-power ultrasonic dissection in the vicinity of these structures. Each generator has five power levels (1–5). Although the vibration frequency is unaltered, the excursion at the tip of the instrument varies from 50 μm at power setting 1 to 100 mm at power level 5 (Figure 2.32). Collateral damage to vital structures is diminished by operating at a maximum power setting of 3. The trade-off is slower cutting, but safety is increased. For coagulation, the power setting should be minimal (level 1) and pressure is applied gently to coapt the vessel walls and create a seal over a wide area before cutting it.

Lasers

Laser is an acronym for 'light amplification by stimulated emission of radiation' and essentially consists of a beam of highly collimated (waves are in synchrony) monochromatic light (of a single wavelength). Visible light is the relatively small range of wavelengths of the electromagnetic spectrum (380–700 nm) that can be detected by the human eye (Figure 2.33). Electromagnetic radiation of wavelengths below (ultraviolet, X-rays and γ-rays) or above (infrared, microwaves and radiowaves) this range is invisible. Laser beams can be generated of wavelengths that cover the entire visible portion of the electromagnetic spectrum (380–700 nm) as well as the ultraviolet and infrared regions.

Production of laser beams

The three components of a laser device are the lasing medium, a power source, often referred to as pump source, and a resonator or optical cavity (Figure 2.34). The lasing medium may be solid (crystals, e.g. ruby, or semiconductor diodes, e.g. gallium arsenide), gas (CO_2, helium, argon, metal vapours) or liquid (usually organic dyes in suitable solvents, e.g. rhodamine 6G in methanol). The lasing medium is energized by the pump source (electric current or light) inside the resonator. This is essentially a vacuum tube with a mirror at either end, one being totally reflective, i.e. bounces back all incident photons (light particles), and the other partially reflective, i.e. lets some photons through, and thus provides the exit point of the laser beam.

Within the resonator, the atoms of the lasing medium are excited from their resting or ground state to a higher energy level by absorption of the electrical or light energy from the power source. When the lasing medium has been excited sufficiently, a stage is reached when more of its atoms are at the upper energy level than at the ground state. This population inversion of the molecules of the lasing medium is essential for the lasing action to begin. Then, some of the excited atoms return spontaneously to the ground state, releasing this extra energy as photons of light of a wavelength characteristic of the lasing medium used (spontaneous emission of radiation). These photons interact with other excited atoms, which then return to the ground state, releasing more photons (stimulated emission of radiation). The net result is the production of waves of the same wavelength (monochromatic) travelling in the same direction and in phase with one another (spatially coherent). These photons are reflected backwards and forwards between the two mirrors, causing further stimulated emission of radiation from other excited molecules of the lasing medium. The partially reflective mirror allows some of this radiation to escape from the resonator, forming the laser beam.

Properties of laser beams

Most lasers produce collimated light; that is, the beam is parallel with minimal divergence with distance. This property enables high irradiance of tissue (power per unit area irradiated or power density). The semiconductor diode-array lasers are an exception as the light output from these solid-state devices is poorly collimated and requires focusing by suitable optics.

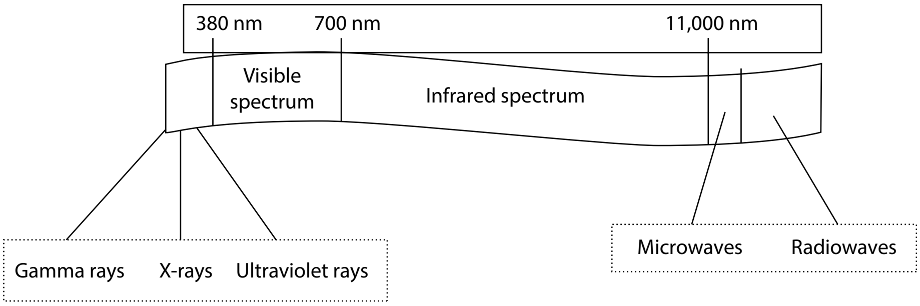


Figure 2.33 The electromagnetic spectrum.

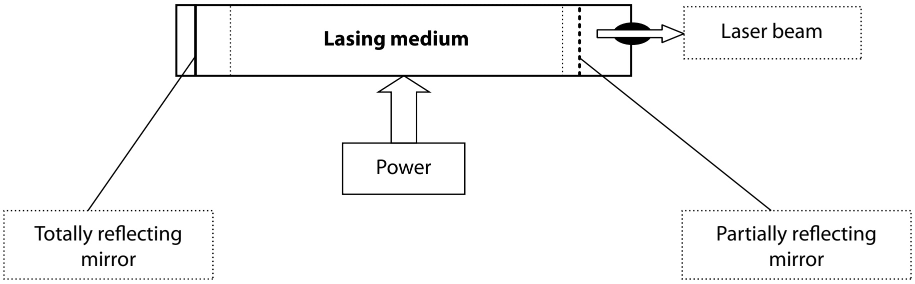


Figure 2.34 Components of laser devices: the lasing medium, a power source, often referred to as pump source, and a resonator or an optical cavity.

The second property is being monochromatic (single or very narrow lines at characteristic wavelengths). Therapeutically, this enables the choice of laser for a given effect, coagulation, cutting or activation of specific chemicals. The third property of laser beams is spatial coherence of the component waves (in phase), although this is reduced to some extent when laser light is transmitted through an optical fibre and is rapidly lost as the light penetrates the tissue.

Classification

Lasers can emit light continuously (continuous-wave lasers) or in the form of discrete multiple pulses of higher energy around 1 ms (pulsed lasers). Both types can be Q-switched. Essentially, this consists of a very fast shutter in the resonator cavity between the lasing medium and partially reflecting mirror. When the shutter is closed, there is a tremendous build-up of light energy in the chamber, which is released as a giant high-energy pulse when the shutter is open (Figure 2.35).

The majority of medical lasers operate at specific wavelengths depending on the lasing medium. Thus, if a different wavelength is needed, another laser system has to be used. This has been the main limitation of lasers in medical therapy. A tunable laser, i.e. one that enables the clinician to select different output wavelength beams for a wide range of specific applications, does not currently exist, although the dye lasers and the more recent solid-state diode-array lasers are tunable over a narrow range. Diode-array lasers will probably replace all other types in view of their portability and because they are largely maintenance free.

From the safety aspects (potential damage to the eye and skin), lasers are classified as 1, 2, 3A, 3B and 4, with class 1 being the safest and 4 the most hazardous, requiring strict safety procedures. The vast majority of medical lasers fall into class 4 because of the power needed. For this reason, laser safety management is mandatory within the UK National Health Service.

Laser safety management

This involves:

- A laser protection adviser: a physicist certified for competence in laser radiation protection.
- Provision of local rules for each specific application of a laser. These address the nature of hazards, training, user's responsibilities, methods of safe working, personal protective equipment, especially eyewear, etc.
- Laser protection supervisor: to ensure that the local rules are implemented and is usually the clinician responsible for the use of the laser.
- Laser controlled area: the operating environment of the laser. The occupancy and activity of all persons in this area are subject to control and supervision to prevent exposure to radiation in excess of the maximum permissible exposure (MPE). This is the level of laser radiation to which the eye and the skin may be exposed without suffering adverse effects. The MPE varies with the wavelength of the radiation, the pulse duration or exposure time and, for lasers operating in the 400–1400 nm range, the size of the retinal image. The distance at which the radiation exposure from the laser equals the corneal MPE is defined as the nominal ocular hazard distance. This is taken into account when specifying the boundaries of the laser controlled area. Warning signs are provided at every entrance to the laser controlled areas that medical lasers that fall within class 3B or 4 should only be operated by authorized users and a register kept of all authorized users for each laser installation.

Examples of medical lasers and their characteristics are shown in Table 2.5.

Medical applications of lasers

Lasers are used to achieve (1) therapeutic effects, (2) stimulation of healing, (3) diagnosis and (4) optical alignment.

Therapeutic effects

These are achieved by the deposition of laser energy at a radiant exposure of 10^5–10^7 J/m^2 and may be (1) photochemical, (2) photothermal, (3) photoablative or (4) photomechanical.

Photochemical effect

In clinical practice, this is referred to as photodynamic therapy (PDT). The basis for the treatment is the prior administration of a photosensitizer, e.g. haematoporphyrin derivative (HpD, Photophrin II) or various chlorin compounds, to the patient.

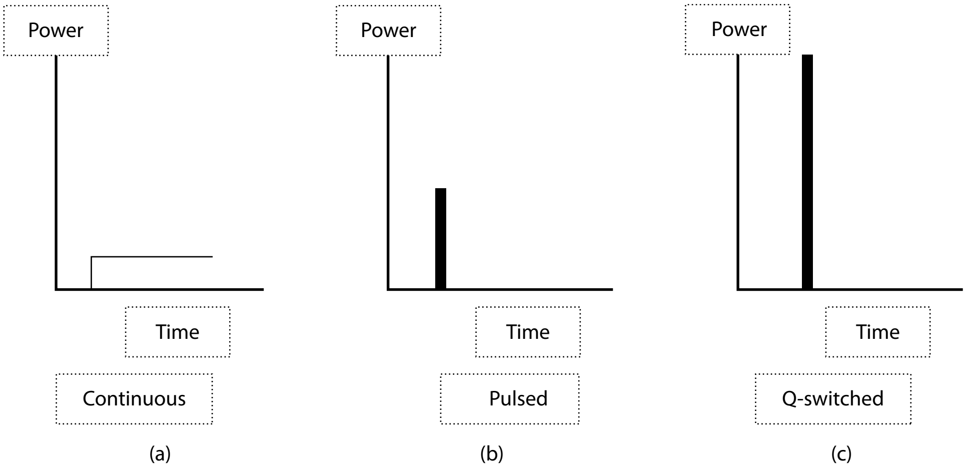


Figure 2.35 Representation of power output of (a) continuous-wave lasers; (b) pulsed lasers; (c) Q-switched lasers. The last releases a giant high-energy pulse when the shutter is open.

Table 2.5 Examples of medical lasers in current usage

Laser type	Wavelength (λ) (nm)	Output power	Beam output	Class 1–4	Beam transport
Helium/neon	630 (red)	0.5–10 mW	CW	1, 2, 3A, 3B	Fibreoptic or mirrors
Argon ion	490–510 (blue–green)	3–10 W	CW and pulsed		
Q-switched	4	Fibreoptic			
Nd:YAG	1060 (infrared)	70 W	CW and pulsed	4	Fibreoptic
Carbon dioxide	10 600	5–30 W	CW and pulsed		
Q-switched	4	Mirrors			
Dye lasers	400–700	15 J	Pulsed	4	Fibreoptic
Excimer	200 (ultraviolet)	0.1 J	Pulsed	4	Direct
KTP	532	15 W	Q-switched	4	Fibreoptic
Diomed diode-array laser	730		4	Fibreoptic	

CW, continuous wave; J, joule (watt-second); KTP, potassium titanyl phosphate; W, watt.

The photosensitizer is taken up and retained by the tumour or dysplastic tissue. If this is irradiated by the appropriate laser light (630 nm in the case of HpD), the light interacts with the photosensitizer (photochemical reaction), leading to the production of highly toxic reactive species such as singlet oxygen with necrosis of the tumour. The actual destruction is far more complex and involves both a direct cytotoxic effect on the tumour cells and destruction of the tumour circulation. With HpD, the major factor in the necrosis of the tumour is destruction of its blood supply. The problem with PDT centres on the prolonged time it takes for the body to clear the photosensitizer. This may vary from several days to weeks (depending on the photosensitizer used). During this time, the patient has to avoid exposure to sunlight because of the risk of a major burn. PDT has been used in palliation, e.g. inoperable oesophageal and bronchial carcinomas, in the ablation of mucosal cancer of the GI tract and urinary bladder, and in the destruction of Barrett's mucosa in association with medical (proton-pump inhibitors) or surgical treatment of the acid reflux. Following this treatment, regeneration with squamous epithelium may occur in some but not all cases.

Photothermal effects

These are used to obtain photocoagulation, cutting and ablation by vaporization. Some 90% of the CO_2 laser radiation is totally absorbed by tissue water within a depth of 0.2 mm. This laser is, therefore, ideal for surgical cutting and for vaporization of tissues. Argon laser penetrates more deeply before absorption (0.5 mm) and neodymium:yttrium aluminium garnet (Nd:YAG; 2 mm). The argon laser is used extensively in ophthalmology and in photocoagulation because of the high absorption of the blue–green light by haemoglobin and melanin. However, for endoscopic control of bleeding the Nd:YAG laser produces more effective coagulation and shrinkage of blood vessels in view of its greater penetration. The addition of a sapphire tip at the end of the conducting fibre or sculptured fibre tips enable contact mode laser delivery, permitting more precise endoscopic application with less smoke than the free beam mode. Interstitial laser hyperthermia produced by implantation of fibres conducting laser light into the tumour is an established modality of *in situ* ablation of tumours in solid organs such as the liver. The pulsed dye laser is the most effective laser for the treatment of port wine stains and has replaced the argon laser for this purpose.

Photoablative effects

Ultraviolet light is highly absorbed in tissue. Excimer lasers (argon fluoride, xenon chloride), which output in the ultraviolet range, are used in ophthalmology for controlled precise ablation, e.g. band keratoplasty and to refashion the corneal curvature for refractive correction. Experimentally, excimer lasers have been used to remove atherosclerotic plaques from peripheral blood vessels.

Photomechanical effects

These require giant pulsed laser energy with power densities of around 10^{16} W/m^2, such as that produced by a Q-switched Nd:YAG laser. These pulses generate localized shock waves and are used in ophthalmology for capsulotomy and iridotomy. Intraluminal stone fragmentation (e.g. impacted ductal calculi) is achieved by creating these shock waves with the end of the optical fibre touching the stones. Other lasers used for intraluminal lithotripsy are the holmium:YAG (2100 nm) and pulsed dye (504 nm) laser systems.

Stimulation of healing

Low-power laser energy (10–100 mW) produced by helium–neon (HeNe), gallium arsenide (GaAs), etc., emitting at the red end of the spectrum or near infrared region, has been shown to stimulate healing and is used in soft-tissue trauma, wound healing and relief of chronic pain due to musculoskeletal disorders.

Diagnosis

The photochemical effect can be used in the early diagnosis of severe dysplasia and *in situ* cancer of mucosal tumours accessible by endoscopy (GI tract, bronchus and urinary bladder). For this purpose, the naturally occurring substance α-amino-levulenic acid (ALA) is administered intravenously or by mouth. ALA is not itself a photosensitizer, but, following administration, it is converted into protoporphyrin ix (Ppix), which fluoresces bright pink when irradiated with violet light (410 nm). The dose given and the power density of the light used are minimal, thereby a photoablative effect is not achieved. The procedure is known as fluorescent endoscopy and has the potential for increasing the earlier detection of severe dysplasia and *in situ* cancer. Initial experience with fluorescent endoscopy using ALA has demonstrated a much higher detection of *in situ* cancer of the oesophagus, bronchus and urinary bladder than ordinary white-light endoscopy. Other photosensitizers can be used for fluorescent endoscopy but ALA is the ideal substance for this purpose because it is rapidly cleared from the body (within 24 hours) and is thus free of the problem of delayed cutaneous sensitivity to sunlight.

Laser Doppler velocimeters incorporating low-power HeNe or diode lasers (up to 2 mW) are in routine use to measure skin blood flow, blood pressure and oxygen saturation (pulse oximeters) during surgery and in high-dependency and intensive care units.

Optical alignment

Low-power lasers emitting in the red are used routinely in optical alignment systems for patient positioning in radiotherapy and also as a light-guide for invisible lasers (Nd:YAG) during photocoagulation.

Cryosurgery

Cryosurgery is a long-established modality for tissue destruction by rapid freezing followed by slow thawing. It has been used in the treatment of skin, head and neck, prostate and liver tumours (primary and secondary cancer). Recent technological developments include laparoscopic probes that enable cryoablation through the laparoscopic approach. The advantage of this is twofold:

1. The procedure can be repeated several times: this is particularly important in the management of patients with secondary hepatic tumours.
2. There is experimental evidence that cryosurgery, carried out through the laparoscopic approach with positive-pressure pneumoperitoneum, results in a greater immediate tumour cell kill. This is probably related to the reduced liver perfusion consequent on the reduction of the portal blood flow by the positive intra-abdominal pressure.

Basis of cryoablation

Rapid cooling of cells to below −40°C results in supercooling and the formation of intracellular ice crystals that disrupt cell membranes and intracellular organelles. In addition, osmotic damage occurs with rapid freezing and slow thawing. During the freezing stage, blood flow ceases completely in the iceball and, when scanned by the contact ultrasound probe, the lesion does not transmit ultrasound waves. These are reflected from the surface of the iceball, giving a characteristic hyperechoic white margin overlying an area of acoustic shadowing (Figure 2.36). As the lesion thaws, blood flow returns for several hours. During this period, fixed tumour antigens, intracellular enzymes, etc., leak into the circulation. The escape of the fixed tumour antigens is thought to initiate an immunological response, although its significance in humans remains uncertain. The serum lactic dehydrogenase level correlates most closely with the extent of cryodestruction on the day after cryosurgery, but transaminases and carcinoembryonic antigens (in the case of secondary deposits from colorectal cancer) are elevated for several days. The revascularization accounts for the 'systemic stress' syndrome after hepatic cryosurgery and limits the extent of freezing possible during any one session (to approximately 12 cm^3). During this period there is a degree of disseminated intravascular coagulation with a fall in platelets, which is maximal between the second and third postoperative day. For this reason, when large volumes of liver are frozen, haematological support is necessary. The revascularization of the cryolesion is only temporary and is invariably followed by occlusion of the blood vessels. The haemorrhagic necrosis becomes encapsulated and undergoes liquefaction over a period of several weeks. Eventually, the lesion shrinks down to a scar.

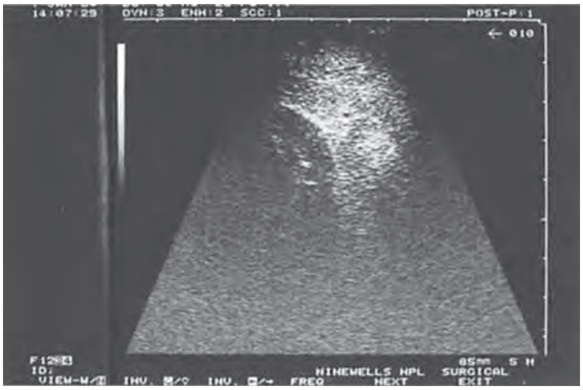


Figure 2.36 Ultrasound appearance of a hepatic iceball during laparoscopic cryotherapy. The ultrasound waves are reflected from the surface of the iceball, giving a characteristic hyperechoic white margin overlying an area of acoustic shadowing.

Modern equipment for cryosurgery of tumours

Although many cooling systems are available, the appropriate coolant used in cryosurgery for *in situ* ablation is liquid nitrogen as this has a boiling temperature of -190°C. All modern cryosurgical machines used for cryoablation of tumours use implantable rather than surface probes. These are shaped like blunt needles and those for laparoscopic use are 2.0 mm and 3.0 mm in diameter. They produce pear-shaped iceballs measuring 4–6 cm in diameter. More than one probe can be inserted into big lesions simultaneously such that the two iceballs merge to encompass the tumour. The objective is to create an iceball that is larger than the lesion by at least 1.0 cm. Modern cryosurgery machines have a gantry into which the cryoprobes dock (Figure 2.37). The cryosurgery process is monitored with ultrasound and thermocouples placed at the periphery of the lesion.

Freezing regimens

Currently, there are two schools of thought. Some practise repeated freeze–thaw cycles, on the basis that the first freeze–thaw cycle alters the thermal conductivity and thus a larger frozen volume is achieved with the second freeze. Other surgeons practise a single freeze, usually of about 20 minutes' duration, on the grounds that there is no material evidence

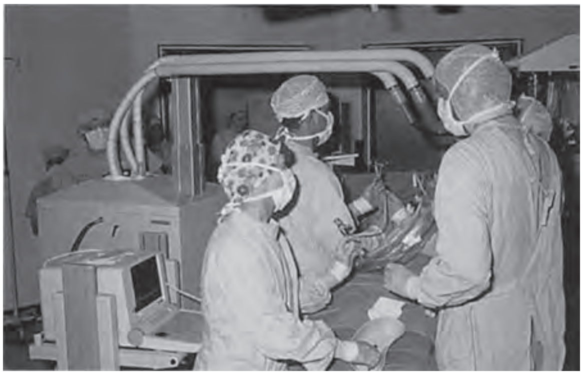


Figure 2.37 Modern cryosurgical unit with a gantry that accepts up to three cryoprobes simultaneously.

that a multiple-freeze cycle regimen produces a larger iceball than a single freeze of the same duration. In addition, the procedure is shorter.

Treatment options

Currently, hepatic cryosurgery is limited to patients who are considered inoperable either because the hepatic reserve is compromised by cirrhosis (primary liver tumours) or because the disease is bilateral and extensive (secondary deposits). However, in the latter instance, cryosurgery may be used in conjunction with resection of a liver lobe. As previously indicated, laparoscopic cryosurgery is preferable to open cryosurgery as it permits repeated application at intervals of 6–8 weeks. Repeated cryotherapy combined with systemic chemotherapy (raltitrexed with high-dose 5-early clincal) is being investigated in phase II trials.

Radiofrequency thermal ablation

In situ thermal ablation of inoperable hepatic tumours is an alternative to cryosurgery since tumour cells are susceptible to temperatures over 45°C. Experimentally, there is a documented synergistic tumour cell kill between hyperthermia and chemotherapy or radiotherapy. Thermal ablation can be achieved by interstitial laser hyperthermia, high-intensity focused ultrasound (HIFU), microwave energy and radiofrequency (RF) current. Interstitial laser hyperthermia requires expensive technology and costly maintenance. HIFU is still at the early clinical stage.

Mechanism of action

RF localized heating is currently the most feasible and cost-effective means of achieving targeted precise thermal ablation and is in established use in cardiology (e.g. treatment of arrhythmias) and neurosurgery. During RF *in situ* tissue thermal ablation, the HF alternating electric current (200 000–500 000 Hz) flows from the non-insulated tip of an electrode into the tissue. This current flow induces ionic agitation within the tissue surrounding the electrode, resulting in frictional heating, which is quite different from electrocautery, in which the current heats the probe. As a result, a localized ellipsoid lesion centred round the non-insulated tip of the probe forms. The thermal lesion is hyperechoic and its generation and size can thus be monitored by ultrasound scanning. The heat generated in the lesion depends on the current intensity (I) and its duration (T) and is described by I^2T. Although current intensity has a much greater effect on lesion volume, a minimum duration of application is needed to achieve the maximum lesion size. During application, heat is continuously generated and lost (by tissue conduction, by convection by the circulating blood and by shunting of RF current along low-resistance electrical paths) until equilibrium, determining the size of the thermal lesion, is reached. Tissue heating decreases rapidly with distance from the tip of the probe ($1/r^4$).

Modern equipment for radiofrequency *in situ* ablation of tumours

The long axis of the ellipsoid thermal lesion is approximately twice the length of the bare tip and its diameter approximates to two-thirds of the long axis. This has meant that until recently the largest thermal lesion that could be achieved *in vivo* by RF thermal ablation was approximately 2.5–3.0 cm and this limited its efficacy. Technological developments in recent years have led to (1) a new generation of microprocessor-controlled RF generators, and (2) multielectrode probes that permit the insertion of an array of single electrodes conducting RF current into the lesion following a single puncture (Figure 2.38). This configuration enables the creation of thermal lesions of up to 4.0 cm in diameter. The new RF generators provide information on the resistance of the tissue surrounding the probe, the temperature of the electrode tip, the RF current intensity passing from the electrode and the RF voltage producing the current. Thus, the control on the size of the lesion is much more precise. A neutral ground pad (electrode) is essential as current travels through the patient. Cooling of the electrode during use improves ablative power.

RF ablation has been mostly used in the treatment of secondary hepatic tumours. It can be applied percutaneously under ultrasound or CT guidance. The laparoscopic approach with contact ultrasound scanning provides a more precise and controlled approach (Figure 2.39).

Other forms of *in situ* ablation

These include high-intensity ultrasound ablation, microwave and nanohyperthermic ablation.

High-intensity ultrasound ablation

Ablation of tumours by HIFU, also known as focused ultrasound surgery, is a totally non-invasive thermal ablation

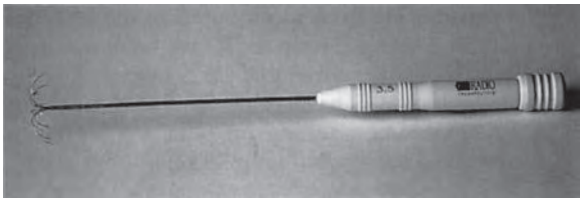


Figure 2.38 Multielectrode probe that permits the insertion of an array of electrodes conducting radiofrequency current into the lesion following a single puncture.

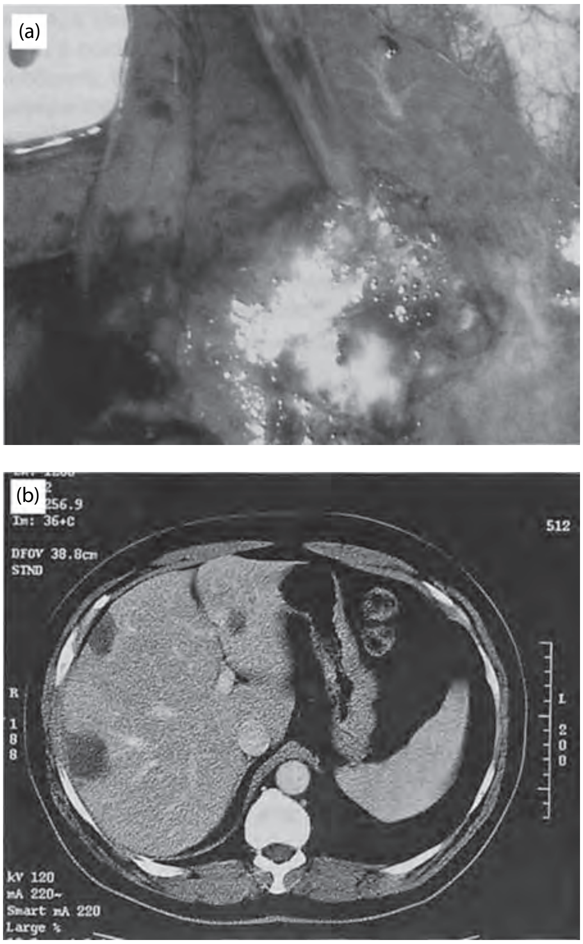


Figure 2.39 (a) Laparoscopic radiofrequency ablation of a secondary hepatic tumour; (b) appearance of the thermally ablated lesion 1 week after treatment: the lesion has a necrotic fluid centre with surrounding fibrosis.

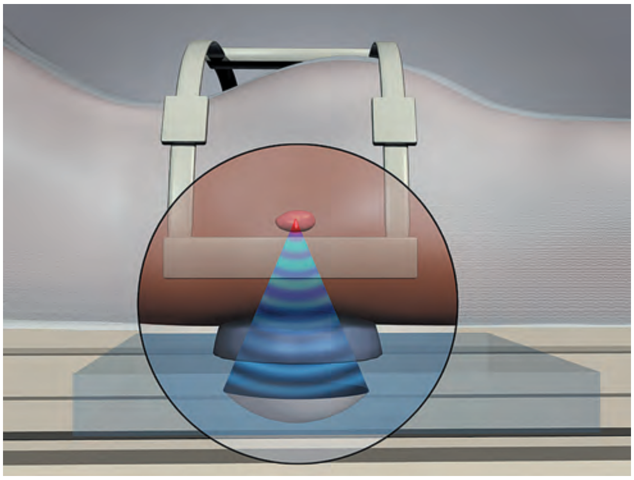


Figure 2.40 High-intensity focused ultrasound (HIFU) totally non-invasive ablation of internal tumours. The transducer is incorporated in a special operating table in the illustration. HIFU probes are also available for ablation of prostate cancer. HIFU systems are guided either by ultrasound (Chinese HIFU system) or by MRI (Insightec system).

of solid tumours. Ultrasound offers two advantages over microwave and radiofrequency techniques for *in situ* ablation: (1) deep penetration through tissues with low attenuation of the beam during transit to the intended ablation site and (2) ability to configure transducers with large dimension (50–100 mm) relative to the wavelength, thereby allowing precise focusing on the ablation zone. The clinical HIFU systems use transducers comprising an array of many elements, each controlled individually for rapid and precise alteration of the beam (Figure 2.40).

There are three forms of HIFU arrays: (1) an annular array comprising 10–20 rings arranged on the surface of a bowl; (2) an annular array divided into eight sectors; and (3) a matrix array with hundreds or thousands of tessellated elements. An annular array can move its focus electronically only along its axis and requires cumbersome mechanical motion for other directions. A sector array additionally allows a small amount of electronic control around the axis, and a matrix array allows full electronic beam control throughout a volume related to the aperture and penetration depth of the array. HIFU arrays have been developed to operate at frequencies from 250 kHz ($\lambda \approx 6$ mm) for transcranial applications to more than 3 MHz ($\lambda < 0.5$ mm) for prostate treatment. Recent developments have enhanced the efficacy of the HIFU system itself but also integrated it with MRI and ultrasound imaging systems for treatment planning and real-time guidance during the ablation.

A major advantage of MRI guidance is that it offers accurate direct temperature measurement for monitoring and control of the hyperthermic ablation in combination with excellent 3D spatial resolution and update rate. *In situ* ultrasound ablation operates mainly by heating and to a lesser extent by cavitation (microbubble formation and implosion). The tumour or pathological tissue is heated rapidly to a temperature above 56°C, often as high as 80°C, causing immediate cell necrosis. One of the advantages of HIFU is the sharply defined nature of the lesion with minimal collateral damage to surrounding normal tissues.

China has more than one HIFU systems manufacturer and has carried out procedures on the largest number of patients, now exceeding 10 000. In the West, HIFU has been used to date in:

- treatment of benign disorders: uterine fibroids, benign prostatic hyperplasia, chronic neuropathic pain
- treatment of small renal carcinomas, especially in renal transplant patients
- treatment of prostate cancer
- palliative care in malignant disease: bone metastases and pancreatic cancer
- primary treatment of solid cancers (phase II clinical trials): breast, kidney, liver, prostate.

Nanohyperthermic ablation

The potential impact on healthcare of nanotechnology is immense and has ushered a new era aptly labelled as ‘nanomedicine’. The EU Technology Platform on Nanomedicine defines nanomedicine as 'the application of nanotechnology to achieve breakthroughs in healthcare'. It exploits the improved and often novel physical, chemical and biological properties of materials at the nanometre scale.

In essence, nanotechnology entails the manufacturing and manipulation of matter at a scale ranging from single atoms to micron-sized objects. From the biological standpoint, nanomaterials match the size of naturally occurring functional units or components of living organisms and, for this reason, enable more effective interaction with biological systems.

Nanoscale and nanomaterials

A nanometre (nm) equals one thousand millionth of a metre and, for comparison, one single human hair measures 80 000 nm across and the diameter of a red blood cell is 7000 nm. By definition, the nanoscale ranges from 100 nm down to the size of atoms (approximately 0.2 nm). At this scale, the properties of materials are very different from their bulk equivalents, i.e. nanomaterials have a relatively larger surface area than bulk material and, for this reason, are more chemically reactive. Additionally, the nanoscale has a marked effect on the strength and electrical properties as the quantum effects dominate the behaviour of materials with respect to their optical, electrical and magnetic properties. Nanomaterials fall into three categories:

1. unidimensional: nanoscale in one dimension, e.g. very thin surface coatings
2. 2D, e.g. nanowires, nanotubes
3. 3D, e.g. nanoparticles, liposomes, buckey balls, etc.

In terms of their complexity and current/future evolution, nanomaterials are classified as passive, active, nanosystems and molecular nanosystems:

- *passive nanomaterials* (first generation): dispersed and contact nanostructures (aerosols, colloids), products incorporating nanostructures (coatings, nanoparticle composites, polymers, ceramics)
- *active nanostructures*: bioactive (targeted drugs, biodevices), physicochemically active (transistors, amplifiers, actuators)
- *nanosystems*: 3D networking and new hierarchical architectures, robotics
- *molecular nanosystems*: molecular devices, machines, biogenerators.

Nanohyperthemic *in situ* ablation

This is still largely experimental but holds great promise for image-guided targeted *in situ* thermal ablation of tumours. The two systems that have attracted considerable research interest both *in vitro* and *in vivo* (animal studies) are carbon nanotubes (CNTs), which may be single or multiwalled and gold nanoshells. One important property of CNTs is their strong absorbance of near infrared light with heating and release of this heat for destruction of cells – nanohyperthermic ablation of tumours. As many solid tumour overexpress surface receptor for folate, CNT-based nanohyperthermic ablation can be made selective by using folate-functionalized CNTs and then producing heating of the tumour by continuous near infrared radiation.

Gold nanoparticles can also be engineered for ablation of tumours and tumour cells most commonly as gold nanoshells, usually coated with polyethylene glycol. They react to near infrared light by the plasmon and photothermal effects. Both techniques are, however, still experimental and it will be some years before they are translated into clinical oncology.

Magnetism

Magnetism is a property by which certain materials respond at an atomic level to an applied magnetic field and are referred to as *magnetic* with *ferromagnetism* exhibited by permanent magnets being the strongest. In turn, permanent magnets produce their own persistent magnetic fields. Some materials assume magnetic behaviour only when exposed to an external magnetic field and lose this when the magnetic field is removed. These materials are called *paramagnetic* and include magnesium, molybdenum, lithium and tantalum. Other materials with very weak magnetic moments are repulsed by magnetic fields and are referred to as *diamagnetic*. Other substances, e.g. copper, aluminium, gases and plastic, are not affected by magnetic fields to any measurable extent and are categorized as *non-magnetic*.

Superparamagnetism is exhibited by certain materials in the nanocrystal scale (1–10 nm) and with temperature below the Curie temperature, when the thermal energy becomes sufficient to change the direction of magnetization of the nanocrystals, when the material assumes paramagnetic behaviour, except that instead of each individual atom being independently influenced by an external magnetic field, the magnetic moment of the entire nanocrystal aligns with the magnetic field. A pertinent example used extensively in MRI as contrast is superparamagnetic iron oxide.

A magnetic domain describes a region within a magnetic material that has uniform magnetization, i.e. where the individual magnetic moments of the atoms are aligned with one another and point in the same direction. When heated above a certain temperature (known as the Curie temperature), ferromagnetic materials undergo a phase transition, and their uniform magnetization within a domain disappears since the magnetic domain structure is responsible for the magnetic behaviour of ferromagnetic materials. The magnetic moment of a magnet is a quantity that determines the force and torque that the magnet can exert on electric currents. Both the magnetic moment and the magnetic field may be considered to be vectors with both magnitude and direction. The direction of the magnetic moment points from the south to north pole of a magnet; and the magnetic field produced by a magnet is proportional to its magnetic moment.

Other useful properties of magnets include:

- *anisotropic magnet*: has a preferred direction of orientation, i.e. magnetic properties are optimal in one direction
- *isotropic magnet*: does not have a preferred direction of magnetic orientation

- *coercive force*: the demagnetizing force necessary to reduce induction *B* to zero from the saturation state of the magnet
- *Curie temperature*: at which material loses its permanent magnetic properties
- *magnetic flux*: total magnetic induction over a given area
- *Gauss*: lines of magnetic flux/cm^2, measured in tesla (SI system)
- *length of air gap* length of the central flux path across the gap
- *remenance*: residual magnetic induction in a magnetic circuit after removal of magnetizing force.

The various types of permanent ferromagnetic materials are:

- *Al–Ni–Co–Fe (alnico)*: anisotropic along one axis – directional properties
- *ferrite*: pressed and fired strontium ferrite; immune to magnetic change with handling and high corrosion resistance
- *Nd–Fe–B*: highest performance magnets but the maximum operating temperature is 100°C; has a low resistance to corrosion but may be coated; anisotropic
- *samarium cobalt* (rare earth cobalt): substantially higher flux densities than alnico and ferrite
- *magnetic sheets*: flexible bonded sheet consisting of fine magnetic powder (barium ferrite) loaded in a flexible thermoplastic binder.

The most commonly used is Nd–Fe–B, followed by samarium cobalt.

Surgical applications

The surgical application involving the use of external magnets includes (1) retraction of hollow viscera to the anterior abdominal wall during laparoscopic surgery and SPLS; (2) external magnetic displacement of laparoscopic intraperitoneal free micro-robots and small CMOS cameras; (3) magnetic instruments for laparoscopic/endoscopic interventions based on prior localized ferromagnetization of tissues needing retraction; (4) magnetic probes, which may be permanent or electromagnetic, the latter with the use of advanced sensor technology can provide both contact and non-contact retraction; and (5) magnetic compression anastomosis.

Ferromagnetization of tissues

Technologies are being developed for localized ferromagnetization of tissues such that these can be manipulated or retracted by magnetic probes. The localized ferromagnetization is achieved in two ways:

1. surface: uses glue-based ferromedia or ferromagnetic mucoadhesive polymeric films
2. injection: special ferrofluids (intramural injection) or glue-based ferromedia (intraluminal injection).

Ferrofluids and ferromagnetic media

Ferrofluids are stable colloidal suspensions of subdomain magnetic nanoparticles (10 nm) coated with a stabilizing dispersing agent (surfactant) that prevents nanoparticle agglomeration even when a strong magnetic field gradient is applied across the ferrofluid. Following injection, even a small 3.0 cm permanent magnet at the tip of a plastic probe can provide excellent contact retraction (Figure 2.41).

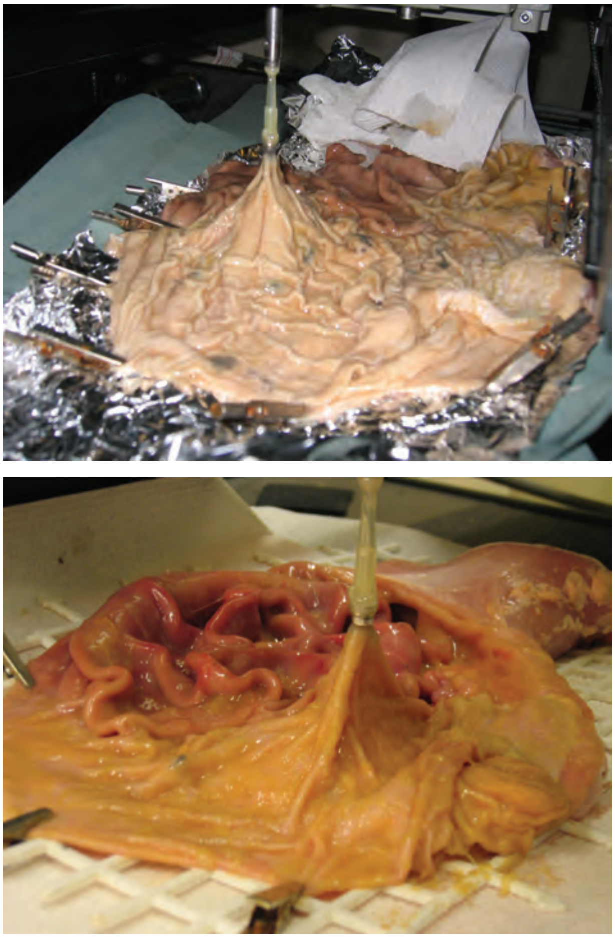


Figure 2.41 (a) Experimental magnetic retraction following submucosal injection of ferrofluid. (b) Mucosal retraction by a polymerized pellet of ferromagnetic glue medium.

Ferromagnetic media are similar suspensions in surgical glues (biological or synthetic) that are either layered on the surface of the organ to be retracted or injected into the bowel lumen. The ferromagnetic glue-based medium (usually cyanoacrylate based), when layered by applicator, binds on polymerization to the surface or serosa and the resultant pellet serves as an attracting point to the magnetic probe.

Magnetic compression anastomoses

The concept of suture less magnetic anastomosis is not new, and various technologies, such as polymer constructs (Valtrac for colonic anastomosis) and much more recently superelastic Ni–Ti (50:50), have been used to achieve intestinal anastomosis. In magnetic compression anastomosis, a magnetic dipole, consisting of two circular or oval magnets (may be ring shaped) with one magnet placed in each of the two hollow organs or intestinal loops to be anastomosed, results in immediate approximation of the two hollow organs as

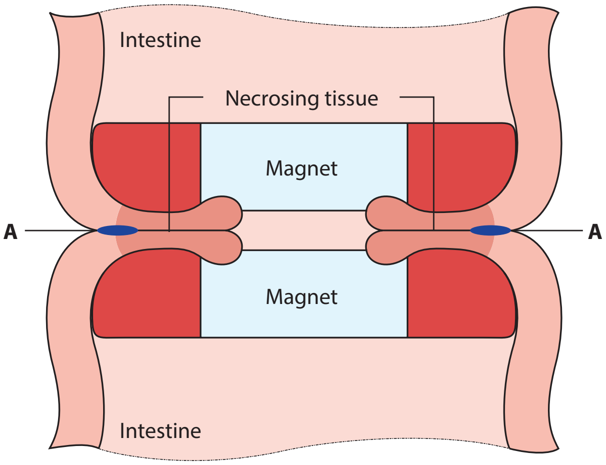


Figure 2.42 Magnetic compression anastomosis.

magnets are attracted to each other forming the dipole. What is different from other types of compression anastomoses is that the dipole exerts uniform compression with pressure necrosis over the area encompassed by the opposing surfaces of the magnets (Figure 2.42). In addition, there is stimulation of fibroblast proliferation with fibrous union between the two walls of the opposed hollow organs surrounding the magnetic dipole (points A in Figure 2.42) over a period of 5–6 days, when usually the magnetic dipole falls through the perfectly circular anastomosis and is eliminated per anus.

Magnetic compression anastomosis is used clinically for end-to-side coronary anastomosis (Ventrica Inc.), colorectal anastomosis (magnetic rings embedded in polyester) and bilioenteric anastomosis. One seminal reported case indicates the immense potential of the technique. The patient developed a stricture of the bilioenteric anastomosis after hepatic transplantation. Under total radiological control, one magnet was inserted in the hepatic duct transhepatically and the other in the jejunal loop distal to the stricture. The patient obtained total relief of jaundice following the magnetic compression hepatojejunostomy without recourse to an operation.

Robotically assisted surgery

Although often referred to as robotic surgery, this term is incorrect as the surgeon operates a sophisticated master–slave manipulator (MSM) system that reproduces the surgeon's hand movements and, through electronic processing, scales down the movements as directed by the surgeon providing great precision and abolishing tremor. In surgery there has been two such operating MSMs: Zeus (Computer Motion) and da Vinci (Intuitive Surgical), but the former is no longer in use. Thus, in essence, over the last 15 years the only system available has been the da Vinci system, although a new Canadian operating MSM (Titon Medical) is due to be launched for phase I clinical examination.

Surgical robots

There are several types of robotic devices used in various clinical scenarios in surgical practice:

- robotic arms with voice control used as camera holders in laparoscopic surgery (Aesop)
- robots used in CyberKnife (gamma-knife) radiotherapy for brain and spinal cord tumours
- robot used in orthopaedic surgery for pinning of fractures to avoid radiation exposure to the surgeon and staff
- RP7 Tele-rounding robot, which enables ward rounds of patients by a specialist from a distance
- operating MSMs for robotic-assisted MAS operations
- robotic systems for single-port/incision laparoscopic surgery
- mini-intracorporeal robots (still experimental).

Master–slave manipulator operating systems

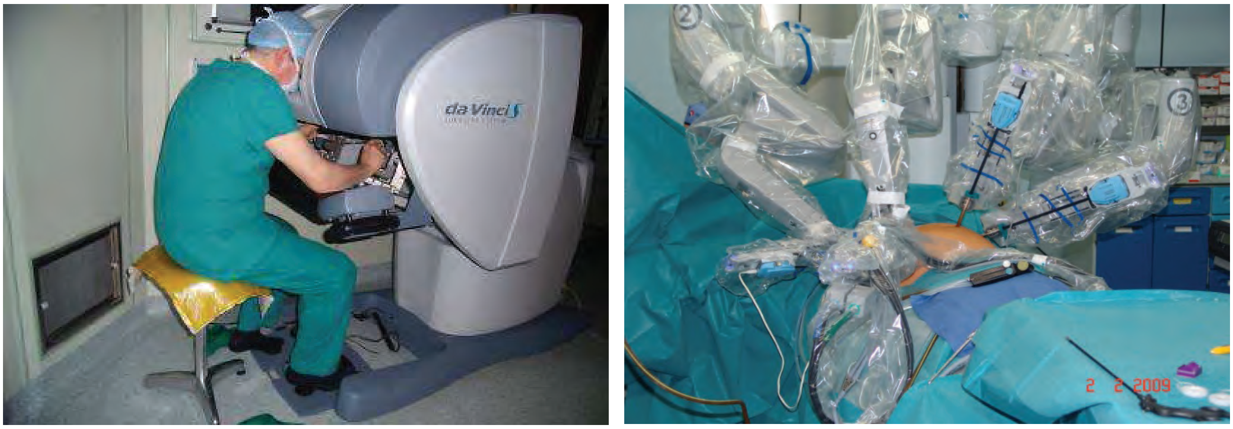
These provide an electronic platform that translates electronically the hand movement of the surgeon via robotic manipulator arms to the movement of the functional tip of the instrument (known as the end-effector) inside the patient and are used largely in complex MAS operations. They greatly facilitate the execution of complex tasks such as intracorporeal suturing by restoring all six DOFs and thereby abolishing the restricted manipulation that the surgeon has to overcome in direct manual laparoscopic surgery. In addition, the system permits motion scaling, whereby the hand movements are faithfully reproduced but scaled down to order. This has two advantages: (1) abolition of transmitted hand tremor of the surgeon and (2) vastly increased precision. Currently, there is only one MSM in well-established use in surgery – the da Vinci (Intuitive Surgical) – as the other competitor (Zeus) is no longer in production or clinical usage. The da Vinci robot has been in clinical use for 20 years and there is a considerable published literature on its use in the various surgical specialties: cardiovascular surgery, general surgery, gynaecology and urology; to date, it has enjoyed a total monopoly, but this is likely to change as several rival systems are due to be launched in 2014 when the patents owned by Intuitive Surgical will run out.

Although often referred to as robotic surgery, this term though commonly applied to these operations is incorrect and the appropriate terminology is robotically assisted surgery.

da Vinci surgical operating robot

The robot was introduced in the USA in 1995, but its first clinical use dates back to 1997 and the first version became available for purchase in 1998. Currently, there are over 2000 installed systems in the USA and another 1000 in other countries. Since 2007, the number of operations performed worldwide with robotic assistance has tripled, increasing from 80 000 surgical procedures to 205 000 cases.

The da Vinci is a teleoperated robotic system based on a master–slave control. It consists of two major subsystems. One subsystem consists of the surgeon's console, housing: (1) the image display, (2) the surgeon's master interfaces,



(a)

(b)

Figure 2.43 (a) Robotic arms covered with sterile plastic drapes with end-effectors (instruments) introduced through ports inside the insufflated peritoneal cavity. (b) Surgeon operating from da Vinci console.

(3) the surgeon's user interface and (4) the electronic controller. The second subsystem is the patient-side cart, consisting of four slave manipulator arms: three for telemanipulation of surgical instruments and a fourth dedicated to the endoscopic camera. The da Vinci system creates an immersive operating environment for the surgeon by providing both high-quality stereoscopic imaging and a man–machine interface that directly connects the surgeon's hands to the motion of the surgical instrument tips. The surgeon visualizes the stereoscopic images by display located above hand level (gaze-down viewing), thereby restoring hand–eye coordination and providing an intuitive correspondence with manipulation. Furthermore, the controller transforms the spatial motion of the instruments into the camera reference frame, such that the surgeon feels as if the hands are inside the patient's body cavity. Lastly, the da Vinci system restores the DOFs lost in direct manual laparoscopic surgery by placing a three-DOF wrist at the distal (internal) end of the instruments, enabling natural wrist pronation/supination, and providing a total of seven DOFs for control of the instrument tip (three orientation, three translation and grip). The system also uses its control system to filter out surgeon tremor, and allows a variable motion scaling from each master to each master–slave arm.

The cart consists of a moveable base with four mounted arms: one for the endoscope and three for instrument manipulation. All four arms are attached to a central column through vertical prismatic joints. Each arm has a set of non-actuated joints that position a distal set of active joints. The passive joints are adjusted manually by releasing the associated brakes. The active joints are controlled by the surgeon.

The surgeon controls the MSM seated on a stool at the computer console, which is positioned remotely from the patient. The console serves as the interface between the surgeon and the robot. The surgeon views the operation through binoculars housed in the console and an infrared beam immediately deactivates the robotic tower whenever the surgeon removes his or her eyes from the binoculars. Several mechanisms permit the surgeon to adjust specific functions of the video system and robotic arms through a panel of buttons on the armrests. The surgeon can also control motion scaling between movements of the masters and the translated motions of the robotic surgical instruments by selecting 1:1, 3:1 or 5:1 motion scaling. The surgeon is also able to control other da Vinci functions with five foot pedals (Figure 2.43a,b).

The da Vinci console provides immersive stereoscopic viewing of the operating field through the binoculars. The imaging is true stereoscopic in nature because of the retinal disparity provided by the optic as distinct from 3D imaging provided by other systems used in laparoscopic surgery, and helps the surgeon considerably in orientation and manipulation within complex operative landscapes.

Robot work space: importance of an optimal port placement

Optimal positioning of the robot and trocar sites is crucial to the expeditious performance of the surgical procedure. Port positioning has to be such as to minimize/avoid collision between the arms of the robot (external to the patient), other obstacles in the operating room and the patient. Other considerations include collision avoidance between the surgical instruments (inside the patient's body) and clashing between the instruments and/or the camera. When work space is large, the ports can maintain an adequate distance between the robotic arms, thereby avoiding external collision when both arms are in active use; in situations of restricted work space, the distance between the ports is reduced, and hence hinders optimal manipulation because of collision.

Advantages offered by the robot

Robotically assisted laparoscopic technology provides a solution to the ergonomic restrictions imposed on the surgeon by direct manual laparoscopic surgery. Several studies have confirmed that the da Vinci robot enhances the surgeon's dexterity as well as provides an ergonomically efficient and user-friendly working environment. Most of the advantages of da Vinci robotic surgery stem from the wristed instrument motions with seven DOFs, scaling for precise movements, elimination of hand tremor, stereoscopic vision and improved ergonomics.

Limitations of the da Vinci robot

Although system malfunctions are well documented in the literature, especially in relation to failures during urological interventions, these are rare and the system is very robust, as demonstrated by a malfunction rate of less than 1%. However, the complete lack of haptic feedback remains a major unresolved problem. The two important adverse consequences of this loss of haptic feedback during laparoscopic robotically assisted surgery are the inability for the surgeon to identify tissue consistency, which enables discrimination between tumour and normal tissue, and the execution of intracorporeal suturing and knot tying, especially with fine suture material, with frequent breakages of the suture material. Although several studies have concluded that operative time is slightly prolonged (because of set-up time), this has been overstated as, when the setting up of the robot is performed by designated and well-trained teams, the increase in operating time becomes insignificant. Other disadvantages of the da Vinci system include its large size which occupies precious space in the OR and restricts access to the patient, distances the surgeon from the patient on the operating table and immerses the surgeon such that he is not aware of what is happening in the OR.

SPRINT robot for single-port laparoscopic surgery

Intuitive Surgical has launched its da Vinci Single-Site kit, and initial clinical results with this system have been reported. However, enhanced dexterity in terms of instrumentation tip manoeuvrability remains unresolved with this kit and, moreover, the large size of the robot and the high cost of the da Vinci will limit its uptake among surgeons. A novel teleoperated robotic platform, SPRINT (Single-Port lapaRoscopy ImaNual roboT), has been developed (Figure 2.44) by a European consortium coordinated by the Pisa group (P. Dario and A. Cuschieri). This is a true multiarm robot aimed at enabling bimanual interventions through a single access port or incision. The manipulator arms of the SPRINT robot are introduced individually through a modified EndoCone into the insufflated peritoneal cavity by means of a cylindrical introducer and then connected to the transverse section. The SPRINT robot is thus assembled internally, and the surgeon is able to control the robot in a master–slave configuration through a dedicated console, allowing the translation of the natural movements of the surgeon's hands to the end-effectors. The SPRINT robot has equivalent functionality and DOF to the da Vinci system.

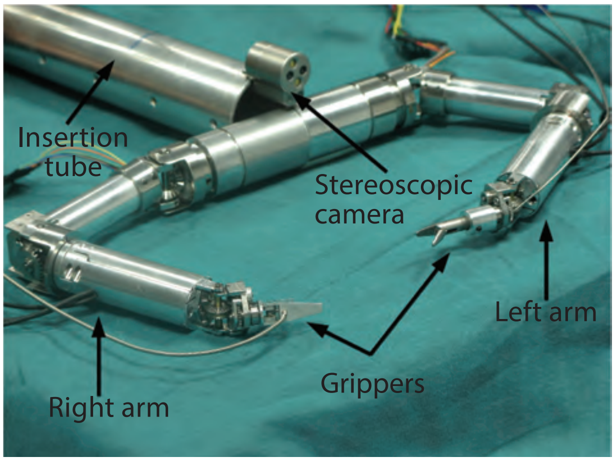


Figure 2.44 Prototype of the SPRINT (Single-Port lapaRoscopy ImaNual roboT) bimanual robot for single-port laparoscopic surgery, which is undergoing laboratory testing at the CRIM laboratory of the Scuola Superiore S'Anna, Pisa, Italy.

Preliminary laboratory tests with the SPRINT robot for SPLS have confirmed its precise functionality and work space necessary for laparoscopic operations, and preclinical animal studies are planned.

GUIDE TO FURTHER READING

American College of Cardiology/American Heart Association Task Force on Practical Guidelines. Guidelines for perioperative cardiovascular evaluation for noncardiac surgery. *Circulation* 1996;**93**:1278–1317.

Barbash GI, Glied SA. New technology and health care costs: the case of robot-assisted surgery. *N Engl J Med* 2010;**363**:701–704.

Becker JM, Dayton MT, Fazio VW, *et al.* Prevention of postoperative abdominal adhesions by a sodium hyaluronate-based bioresorbable membrane: a prospective randomized double-blind multicentre study. *J Am Coll Surg* 1996;**183**:297–306.

Cherukuri P, Glazer ES, Curley SA. Targeted hyperthermia using metal nanoparticles. *Adv Drug Deliv Rev* 2010;**62**:339–345.

Christensen CM. *The Innovator's Dilemma.* Boston, MA: Harvard Business School Press, 1997.

Cuschieri A, Hanna GB. Surgical competence, training and privileges. In: Hobsley M, Treasure T, Northover J (eds) *Laparoscopic Surgery.* London, UK: Arnold, 1997:21–41.

Cuschieri A, Wilson RG, Sunderland G, *et al.* Training initiative list scheme (TILS) for minimal access therapy: the MATTUS experience. *J R Coll Surg Edin* 1997;**42**:295–302.

Emerich DF, Halberstadt C, Thanos C. Role of nanobiotechnology in cell-based nanomedicine: a concise review. *J Biomed Nanotechnol* 2007;**3**:235–244.

Gannon CJ, Cherukuri P, Yakobson BI, *et al.* Carbon nanotube-enhanced thermal destruction of cancer cells in a noninvasive radiofrequency field. *Cancer* 2007;**110**:2654–2655.

Hensman C, Baty D, Willis RG, Cuschieri A. Chemical composition of smoke produced by high-frequency electrosurgery in a closed gaseous environment. *Surg Endosc* 1998;**12**:1017–1019.

Kroh M, El-Hayek K, Rosenblatt S, *et al.* First human surgery with a novel single port robotic system: cholecystectomy using the da Vinci Single-Site platform. *Surg Endosc* 2011;**25**:3566–3573.

Morel P, Hagen ME, Bucher P, *et al.* Robotic single-port cholecystectomy using a new platform: initial clinical experience. *J Gastrointest Surg* 2011;**15**:2182–2186.

Muller W, Fritzsch G. Mechanicotechnical basis of surgery using invasive ultrasound energy. *Endosc Surg Allied Technol* 1994;**2**:205–210.

Petroni G, Niccolini M, Menciassi A, *et al*. A novel robotic system for single-port laparoscopic surgery. *Surg Endosc* 2013; in press.

Tang B, Hou S, Cuschieri A. Ergonomics of and technologies for single-port laparoscopic surgery. *Minim Invasive Ther Allied Technol* 2012;**21**:46–54.

Turchetti G, Palla I, Pierotti F, Cuschieri A. Economic evaluation of da Vinci-assisted robotic surgery: a systematic review. *Surg Endosc* 2012;**26**:598–606.

CHAPTER 3

Risk assessment and prognostic indices in surgery

OMAR FAIZ

Introduction

> Take calculated risks. That is quite different from being rash.
> George S. Patton, United States General (1885–1945)

Risk assessment is a fundamental component of contemporary surgical practice. It is important for three reasons. First, appropriate selection of surgical patients depends upon the balance between the benefit likely to be derived from the surgical procedure and the risk posed by that intervention. Second, an assessment of risk will guide the surgeon and anaesthetist as to the degree of supportive care that will be required in the postoperative period. Sometimes risk assessment will also identify factors amenable to optimization in the immediate preoperative phase. Last, adjustment for surgical risk is essential for accurate performance evaluation. Increasingly, surgeons' outcomes are subjected to professional and public scrutiny. Adjustment of outcome measures for risks that are outside the surgeon's control are essential for meaningful representation of morbidity and mortality rates in this context.

Risk assessment can take various forms. At its most subjective, the surgeon's perception, or 'gut feeling', may offer a crude prediction of outcome. This method is, however, highly dependent on the individual's experience and perception of risk. In order to overcome subjective variation in risk evaluation, quantification is necessary. A number of risk-scoring systems have been developed. Many of these combine patient demographic, comorbidity and physiological parameters along with operative factors to quantify risk. There are many different disease-specific risk assessment scores, and a smaller number of generic scoring systems.

This chapter comprises four sections. In the first, the important risk factors that influence outcome in surgery are described. In the second, methods of preoperative assessment are considered, and the third section describes some of the commonly used risk-scoring systems The fourth section describes methods of patient optimization and risk reduction in surgery.

Risk factors

Risk factors relate to age, comorbid disease, the nature of the operation, the surgeon and the patient's medication:

- age
- comorbidity
  - cardiovascular disease
  - respiratory disease and smoking
  - gastrointestinal: malnutrition, adhesions, jaundice
  - renal disease
  - haematological conditions
  - obesity
  - diabetes mellitus
  - drugs
- surgeon, operative severity and mode of admission (i.e. emergency surgery)
- gender
- social deprivation.

Age

Older patients are at increased risk of death and serious complications following major surgery. Most surgeons would agree that no patient should be denied an operation on the basis of age alone. An understanding of how the risks and benefits of surgical intervention change with advancing age is however important for informed decision-making. In the elderly population limited mobility, frequent presence of intercurrent disease and diminished reserve of cardiac, respiratory or renal function may potentially impede recovery.

It has been demonstrated in many differing specialties that outcome is inferior following surgery in older patients when compared with results from younger counterparts. For example, wound infection rates tend to be higher in elderly patients. Furthermore, postoperative stroke occurs in 1% of patients above the age 65 years and in 3% of those above 80 years. There does not appear to be an increased incidence of postoperative stroke in patients with a history of old stroke or in those with a symptomatic carotid bruit. However, a recent stroke (within 2–3 months of the operation) is thought to increase the risk of a recurrent episode. As the autoregulation of the cerebral circulation becomes impaired with age, extremes of hypotension and hypertension must be avoided throughout the perioperative period.

It is debated whether the increased risk of mortality presented to older patients undergoing major surgery is dependent on chronological age *per se* or rather that this cohort are more likely to have associated comorbidities. Certainly, elderly patients undergoing major planned surgery for colorectal cancer are more likely to die following surgery than younger patients with a similar comorbidity burden. In addition, their risk of mortality in the months that follow surgery is significantly higher than that of age-matched population controls. These mortality risks are even greater when patients present as emergencies. Patients aged older than 80 years who undergo emergency surgery for large bowel obstruction have a 30 day mortality rate that exceeds 20%.

Elderly patients therefore require careful preoperative evaluation to assess their cardiac, respiratory and renal function. An understanding of the patient's exercise tolerance, social circumstances and cognitive function are also essential. Increasingly, a multidisciplinary approach, involving surgeons, anaesthetists, elderly care physicians and physiotherapists, is being adopted to optimize perioperative care and rehabilitation in this potentially vulnerable cohort. Commonly undertaken ward interventions, such as intravenous fluid administration and blood transfusion, offer potential risks of circulatory overload in patients in whom cardiac reserve is often limited. Elderly patients also require smaller doses of narcotic analgesics or sedatives, as these are likely to precipitate confusion, both as a result of a direct action on the central nervous system (CNS) and from resulting hypoxia as a consequence of respiratory depression.

Future research will need to identify whether risk assessment in the elderly can be extrapolated beyond mortality measures to include patient-centred outcomes. Factors such as loss of independent living or quality of life that can be brought about through surgery may represent an important decision factor for many patients. In addition, the role of perioperative interventions such as prehabilitation, minimal access techniques (that incur less tissue trauma) and multimodal rehabilitation packages require further study.

Comorbidity

Cardiovascular disease

Overall, myocardial infarction and congestive heart failure are the most common causes of perioperative death in patients undergoing non-cardiac surgery. Lee and colleagues studied over 4000 patients at a tertiary centre undergoing elective non-cardiac surgery. They identified that major cardiac complications occurred in 2% of patients. Six factors were identified as independent predictors of cardiac complications: high-risk surgery, prior history of ischaemic heart disease, congestive cardiac failure, prior history of cerebrovascular disease, preoperative insulin treatment and an elevated creatinine. These factors collectively represent the Revised Cardiac Index. Rates of major cardiac complication occurred at 0.4%, 0.9%, 7% and 11% when 0, 1, 2, 3 or more factors were present respectively.

The Committee of Perioperative Cardiovascular Evaluation for Non-cardiac Surgery [American College of Cardiology (ACC) and American Heart Association (AHA) Task Force] has provided a framework for considering the management of patients with cardiac risk undergoing non-cardiac surgery. The guidelines identify three broad categories, including major, intermediate and minor clinical predictors of increased perioperative cardiovascular risks such as myocardial infarction, congestive cardiac failure and death (Box 3.1). When present, major predictors necessitate intensive management and delay or cancellation unless the condition requires emergency surgery to save life. Intermediate predictors enhance the risk of perioperative cardiac complications and require careful assessment by objective testing of the patient's cardiac performance. It should be noted that a history of myocardial infarction (occurring more than 30 days previously) or pathological Q waves on the ECG are considered as intermediate risk factors, compared with a recent myocardial infarction, which is listed as a major predictor. Minor predictors are recognized markers for cardiovascular disease that have not been proven independently to increase perioperative risk.

BOX 3.1 Clinical predictors of increased perioperative cardiovascular risk (myocardial infarction, congestive heart failure, death)

Major:
- Unstable coronary syndromes
  - Recent myocardial infarction with evidence of important ischaemic risk by clinical symptoms or non-invasive study
  - Unstable or severe angina (Canadian class III/IV)
- Decompensated congestive heart failure
- Significant arrhythmias
  - High-grade atrioventricular block
  - Symptomatic ventricular arrhythmias in the presence of underlying heart disease
  - Supraventricular arrhythmias with uncontrolled ventricular rate
- Severe valvular disease

Intermediate:
- Mild angina pectoris (Canadian class I/II)
- Prior myocardial infarction by history or pathological Q waves
- Compensated or prior congestive heart failure
- Diabetes mellitus

Minor:
- Advanced age
- Abnormal ECG (left ventricular hypertrophy, left bundle branch block, ST–T abnormalities)
- Rhythm other than sinus (e.g. atrial fibrillation)
- Low functional capacity (inability to climb one flight of stairs)
- History of stroke
- Uncontrolled systemic hypertension

In practice, these guidelines translate into the following management pathway:

- *Major predictors present*: surgery cancelled unless the condition is immediately life threatening. Intensive cardiological management takes priority. Some may require myocardial revascularization (when the stress of elective non-cardiac surgery is likely to exceed the stress of daily life) in the first instance.
- *Intermediate predictors present*: objective performance of cardiac status (e.g. echocardiography, stress ECG) undertaken, on which is based the individual risk that should influence the decision to proceed with surgery.
- *Minor predictors present*: surgery can proceed with adequate monitoring and postoperative support.

Although moderate hypertension is not an independent risk factor, its presence is often an indictor of coexisting coronary artery disease (CAD) and intraoperative fluctuations of the blood pressure in hypertensive patients are associated with myocardial ischaemic episodes. Furthermore, patients with preoperative hypertension are more likely to develop hypotension during surgery than normotensive patients, especially when blood volume is decreased. Thus, hypertension should be controlled prior to elective surgery. The exact details of any antihypertensive therapy that the patient is taking must be recorded, so that the anaesthetist can avoid adverse drug interactions and hypotensive episodes. Dysrhythymias are also important and patients with heart rates of less than 40 beats/min should be considered for pacemaker insertion preoperatively, even as an emergency. Likewise, patients with Stokes–Adams attacks have a great risk of developing complete heart block during anaesthesia and should also be paced preoperatively.

Respiratory disease and smoking

The incidence of respiratory disease in surgical patients varies with the population and in Western countries this ranges from 25% to 50%. The presence of either obstructive or restrictive pulmonary disease increases the risk of postoperative pulmonary complications and carries an adverse effect on the cardiovascular system. Thus, hypoxaemia, hypercapnia, acidosis and the increased work of breathing can cause further deterioration in patients with compromised cardiac function. Chronic obstructive airways disease is followed by a higher incidence of postoperative pulmonary complications than is restrictive airway disease. In patients with significant pulmonary disease, estimation of functional capacity (spirometry and blood gas analysis) forms part of the preoperative evaluation. Wherever possible, surgery in patients with acute respiratory infections should be postponed until 2 weeks after the resolution of the infection. The timing of elective surgery in patients with chronic respiratory disease should coincide with remission and after a period of intensive physiotherapy and appropriate treatment with antibiotics and bronchodilators. Postoperative problems, particularly chest infections and segmental or lobar collapse, are exceedingly common in patients with chronic bronchitis and in heavy smokers.

Smoking increases the risk of surgery and anaesthesia as a result of its adverse effects on the cardiovascular and respiratory systems. Carbon monoxide and nicotine are responsible for the immediate cardiovascular effects. Consequent on the formation of carboxyhaemoglobin, carbon monoxide reduces the amount of haemoglobin available for combination with oxygen and alters the oxygen dissociation curve such that the affinity of haemoglobin for oxygen is enhanced. It also has a weak negative inotropic action on the heart. Nicotine causes an increase in heart rate and blood pressure; thus, it enhances the demand of the myocardium for oxygen while carbon monoxide decreases the supply. Elimination of both carbon monoxide and nicotine with improvement in the cardiovascular fitness is complete following a 12–24 hour abstention from smoking.

There is a sixfold increase in the postoperative respiratory morbidity among patients who smoke more than 10 cigarettes per day. The responsible factors include small airways disease, hypersecretion of a thick viscid mucus and impairment of tracheobronchial clearance. Smoking also depresses the immune system. It induces reduction in immunoglobulin levels, natural killer cell activity, neutrophil chemotaxis and pulmonary alveolar macrophage activity. It is therefore recommended that smokers should abstain from smoking for about 3 months prior to surgery, as this will result in an improvement in pulmonary function, a reduction in the likelihood of postoperative respiratory morbidity and the return towards a normal immune response. Those who find it impossible to stop smoking for this period will derive some benefit in terms of improved cardiovascular function from a short period of abstinence (12–24 hours before their operation).

Gastrointestinal

Malnutrition

The malnourished patient is at increased risk of postoperative morbidity. However, assessment of nutritional status can be difficult. In essence, it is important to obtain an accurate history. There is evidence that loss of 15–20% of body weight is likely to be associated with impairment of physiological function and increased postoperative complications. Physical examination should be directed towards determining body stores of fat and protein. Inspection of the temporalis, spinatus and interosseous muscles is a good indication of muscle wasting and the presence of oedema is suggestive of protein (kwashiorkor-like) malnutrition. Contrary to popular belief, the level of plasma albumin is not a particularly good index of malnutrition. Albumin is most useful as a negative acute phase reactant and in a patient who is metabolically stressed, usually by sepsis, a low serum albumin can be expected. Thus, in the patient who is suspected of malnutrition and has a low plasma albumin, a focus of sepsis should be sought.

In malnourished patients with benign disease in whom surgery may be delayed it is common practice to attempt a period of supplemental nutrition prior to surgery. As a general guideline, body weight loss of 15–20% below the patient's ideal weight, accompanied by clinical evidence of muscle weakness and wasting, is still a reasonable indication for nutritional repletion before undertaking surgery. If possible, this should be carried out using the enteral route, thereby preserving gut immunological function. In patients with severe gastrointestinal malfunction parenteral nutrition may be required.

Adhesions

Another important gastrointestinal problem, which may affect the outcome of surgery, is the presence of intra-abdominal adhesions. Postoperative adhesions are dealt with elsewhere. Previous laparoscopy is unlikely to be associated with significant adhesions, but in the patient who has undergone a laparotomy there is a high likelihood of intra-abdominal adhesion formation. Reoperative surgery is usually lengthier and associated with greater blood loss than primary laparotomy. When multiple dense adhesions are found within the abdominal cavity, great care should be exercised in dissection as accidental enterotomies lead to a significant risk of postoperative intestinal fistulation.

Jaundice

The patient with jaundice is at significant risk of postoperative complications, particularly sepsis, disorders of clotting, renal failure, liver failure, and fluid and electrolyte abnormalities. In addition, the conjugation and metabolism of drugs and anaesthetic agents are impaired secondary to hepatocyte malfunction.

Contrary to popular belief, there is no evidence to support the view that wound healing is impaired in the presence of jaundice. Wound-healing problems are usually confined to patients who have jaundice as a result of underlying malignancy and are the result of the disease and its association with poor nutrition.

Nutritional deficits in jaundiced patients are variable and parenteral nutrition should only be used very selectively because of the risk of infection. A high intake of carbohydrate is essential and amino acid solutions containing aromatic amino acids should be used sparingly as these may precipitate encephalopathy in susceptible patients. The oral route for nutritional supplementation should be used whenever possible.

Infective complications are usually due to the fact that most patients with jaundice undergoing surgery have an obstructed biliary tree and the bile in these patients may be infected by aerobic Gram-negative organisms. Coagulation disorders are usually due to a prolonged prothrombin time resulting from a deficiency of vitamin K-dependent factors consequent on the malabsorption of this vitamin that occurs in cholestatic jaundice. Under normal circumstances the patient with jaundice who is about to undergo surgery merely requires an intramuscular injection of vitamin K. This will reverse the multifactorial clotting deficiency over 1–3 days. With severe hepatic disease, the prothrombin time may remain abnormally prolonged despite this treatment and in this case administration of fresh-frozen plasma is necessary to cover the perioperative period.

Renal

It is important to obtain an assessment of renal function prior to surgery. The patient with renal failure may require haemofiltration throughout the perioperative period and it is particularly important to establish acid–base as well as the potassium levels as acidosis and hypo- or hyperkalaemia increase the risk of cardiac dysrhythmias. Renal failure is commonly associated with heart disease and complicates its management. Correction of prerenal failure is particularly important in the emergency situation.

Haematological conditions

Anaemia

Preoperative anaemia multiplies mortality risk in patients undergoing major surgery with other associated renal and cardiac illness. The patient with profound anaemia should have this corrected preoperatively as the reduced oxygen-carrying capacity of the blood can put undue strain on the cardiovascular system as well as increasing the likelihood of chest infection and confusion. Allogenic blood replacement carries risks of transfusion-transmitted infections although these risks are substantially lower than they were three decades ago. In patients with iron-deficient anaemia, this should ideally be corrected by means of oral iron therapy if asymptomatic. A range of preoperative options including intravenous iron therapy, erythropoietin, cell salvage, acute normovolaemic haemodilution and autologous blood transfusion now exist when patients are asymptomatic or mildly symptomatic and there is sufficient time to achieve restoration. Most anaesthetists would consider preoperative allogenic blood transfusion in patients with a haemoglobin level less than 8 g/dL, with a lowered threshold when cardiac comorbidity coexists. Ideally, any patient requiring a blood transfusion should have this performed at least 1–2 weeks prior to surgery for elective conditions to allow for haemodynamic stabilization and not disadvantage oncological outcome. In addition, it must be emphasized, however, that correction of mild anaemia is neither necessary nor desirable perioperatively. In fact, some observational and randomized studies suggest a disadvantage to liberal transfusion in the perioperative period. It is, however, difficult to establish whether the poor outcome identified in the latter studies was attributable to anaemia or as a consequence of surgical blood loss *per se*. Most patients in this category have adjusted to the reduced haemoglobin load and the associated haemodilution may be beneficial in ensuring adequate tissue perfusion in the perioperative period. Indeed, controlled haemodilution down to a packed cell volume of 30–35% has been reported to reduce morbidity.

Polycythaemia, thrombocytosis and other conditions that increase blood viscosity enhance the risk of thromboembolism or haemorrhage or both. Sickle cell disease must also be recognized before surgery. The condition can occur as the homozygous state (SS), resulting in the full-blown disease with attacks of sickle cell crises and haemolytic anaemia, or as the heterozygous condition (sickle cell trait). This rarely causes problems and is associated with a normal lifespan. Although the abnormal haemoglobin S is soluble when oxygenated, it polymerizes and crystallizes out when it loses oxygen. As a result of this conformational change, the red cells become elongated, sickle shaped and rigid, leading to increased blood viscosity and an abnormal rheological state. The rigid cells block the microcirculation to various organs, inducing episodes of pain and infarction (joints, small bones of the hands and feet, spleen, etc.). Attacks of vaso-occlusive crises are precipitated by dehydration, infection, hypoxaemia (during surgery and anaesthesia), severe physical exertion, childbirth and high altitude. Repeated attacks of splenic infarction can lead to an increased susceptibility to infection by encapsulated bacteria. If sickle cell disease is suspected, the patient must be screened and special anaesthetic and perioperative precautions taken if surgery and general anaesthesia are required (adequate hydration, avoidance of hypoxaemic episodes, oxygen therapy).

In patients with a suspected bleeding disorder, a full coagulation screen must be carried out.

Diabetes mellitus

Diabetes is associated with increased need for surgery as well as enhanced perioperative risk. Patients with diabetes mellitus require special management**.** An important feature in the management of the diabetic patient is to avoid dehydration and perioperative instability of blood glucose levels. Patients with diabetes mellitus are predisposed to CAD and, when present, myocardial ischaemia is more likely to be silent than in the general population.

Obesity

Obese patients have an increased risk of respiratory complications, deep vein thrombosis, wound infection and wound dehiscence. In addition, they have a higher incidence of intercurrent disease and restricted mobility. The technical difficulty of the operative procedure is also increased, making the risks of failure of attempted minimally invasive techniques and iatrogenic injury during surgery more likely. Whenever possible, controlled weight reduction is recommended before elective surgical treatment. A significant proportion of obese patients have sleep apnoea, and are at risk of postoperative pulmonary failure. Some patients may require pre-empted respiratory support on the intensive care unit (ICU) perioperatively.

Drugs

Many patients coming to surgery are on prescription medication. Some drugs (e.g. antihypertensives) should be continued right up to the time of operation. Oral contraceptives pose the specific risk of postoperative deep vein thrombosis. In women using the combined pill (oestrogen plus progesterone), the risk of thromboembolic complications is double that of non-users. This is related to a reduction in the activity of antithrombin III induced by the additive affect of the contraceptive and general anaesthesia. This enhanced risk is not seen with the progesterone-only pill, which need not be stopped over the time of elective surgery. Current guidance suggests that oestrogen-containing contraceptives or hormone replacement therapy should be discontinued 4 weeks before major elective surgery and alternative contraceptive arrangements made. In the emergency situation, prophylactic low-dose heparin and graduated compression stockings should be used and early mobilization encouraged. Patients on warfarin are at risk of haemorrhagic complications. These patients should have the warfarin discontinued at least 3–7 days prior to surgery if the risk of venous thromboembolism is low. Surgery can be conducted safely once the international normalized ratio is <1.5. If anticoagulation is considered vital, the warfarin should be converted to a low-molecular-weight heparin or unfractionated intravenous heparin. This can be continued up to 4–6 hours before operation and recommenced about 12 hours postoperatively. The American College of Chest Physicians has issued guidance on antithrombotic prophylaxis in specific circumstances. In some emergencies acute reversal of warfarin is necessary. Under such circumstances liaison with haematologists and administration of reversal agents (vitamin K, prothrombin concentrate complex and fresh-frozen plasma) is necessary. In patients with bare metal or drug-eluting coronary stents requiring surgery within 6 weeks or 12 months of placement, respectively, it is recommended that antiplatelet therapy (aspirin and clopidogrel) is continued throughout the perioperative period.

Operative severity and operating surgeon

The magnitude of an operation as well as whether it is carried out on an elective or expedient basis represent important risk factors for subsequent outcome. Various schedules of operations exist. Operations are often classified into four grades as minor, intermediate, major and major+. Examples of such grades include:

- *minor*: local anaesthetic procedures, excision of skin lesion, drainage of breast abscess
- *moderate*: uncomplicated groin hernia repairs, uncomplicated varicose vein operations, tonsillectomy
- *major*: total abdominal hysterectomy, cholecystectomy with choledochotomy, major amputation
- *major+*: any aortic procedure, total joint replacement, colorectal resection, pancreatic or liver resection, oesophagectomy.

These classifications are used in the operative severity scoring systems described below.

Several studies comparing postoperative mortality and cardiac complications in elective versus emergency operations have concluded that mortality risk and cardiac complications are two to five times more likely to occur following emergency surgical procedures than following elective operations. Emergency presentation is often accompanied by substantial physiological and biochemical derangement. Thus, emergency surgery is itself an important risk factor across a wide range of differing surgeries.

It is also an important, if unpalatable, fact that the individual surgeon is a significant risk factor. It is now generally recognized that the outcome of similar operations can vary widely according to the skill of the operating surgical teams and it is important for all surgeons to be aware of their own results and to ensure that they conform to a reasonable standard. In the future, training must ensure that all surgery is carried out to an acceptable standard.

Preoperative assessment

Preoperative assessment is an essential aspect of surgical care. Eventual outcome will be largely dependent on measures taken as a result of the assessment of operative risks and fitness for general anaesthesia. Preoperative patient evaluation offers an opportunity for

- assessment of operative risks; appropriate patient selection
- assessment of fitness for general anaesthesia and tolerance of indicated surgery
- adequate explanation to the patient of the nature of the operative procedure, so that fully informed consent can be obtained
- optimization of nutritional, blood volume, fluid and electrolyte deficiencies when possible
- institution of prophylactic measures against common postoperative complications
- general preparation of the patient for surgery
- estimation of the amount of blood required to cover the operation
- assessment of the likely postoperative course and probable need for high-dependency or intensive care after the operation.

Careful patient selection involves balancing the relative benefits from a given surgical procedure against the known risks and complications. This decision is taken against the background knowledge of the natural history of the untreated disease from which the patient is suffering as well as the life expectancy and estimated quality of life gains achievable via operative or non-operative courses. In some circumstances a consensual decision is obvious to both patient and clinician. At other times the most appropriate course of action is less clear. Certainly, in the elective setting there is a greater amount of time to allow full consideration. A historic paternalistic approach to decision-making, in which clinicians decide the best option for their patients, is being replaced by shared decision-making models in which mutual agreement is achieved.

Good selection of patients for surgery also entails an early decision that the medical or conservative management has failed since, other risk factors being equal, the overall operative mortality is lower for procedures undertaken under elective conditions. Thus, for example, the mortality following colectomy for ulcerative colitis is highest when this is performed as an emergency because of colonic perforation, intermediate when undertaken urgently for toxic megacolon and lowest when the procedure is performed electively because of failure of medical treatment.

Another aspect of patient selection is referral for specialized treatment. The impact of surgeon case load and subspecialization has been shown across a broad range of surgical interventions. Improved outcome has been demonstrated in patients undergoing oesophagectomy, gastrectomy, thoracotomy, proctectomy and ileoanal pouch when undertaken by high-volume surgeons with appropriate expertise.

History

When assessing a patient for surgery a comprehensive history must be taken. Particular attention must be paid to the following:

- cardiac and vascular disease, including deep vein thrombosis; recent myocardial infarction is a significant risk factor
- respiratory disease and smoking habits
- other medical disorders, particularly hypertension, diabetes, bleeding diatheses and previous stroke
- detailed account of all drugs being taken
- alcohol intake
- previous experience of anaesthetics, especially intractable vomiting, volatile agents used and specific anaesthetic complications, apnoea and malignant hyperpyrexia.

Physical examination

Again, a full physical examination must be carried out. The following are particularly important:

- cardiovascular system
- respiratory system
- nutritional status, body mass index (BMI)
- mental state
- abnormalities of the face, jaw and neck
- presence of dentures.

Investigations

Investigation can be divided into routine preoperative investigations and special investigations.

Routine investigations

When carrying out routine preoperative investigations it is important to ensure that adequate information is obtained, but not at the expense of carrying out large numbers of unnecessary investigations. For this reason it is useful to think of preoperative investigations as mandatory, discretionary or unnecessary in terms of the type of operation or the status of the patient. National guidance is available in the UK for the use of preoperative tests. The requirement for individual tests is based upon the grade of surgery undertaken in addition to American Society of Anesthesiologists (ASA) grade (see below). The standard (or commonly utilized) investigations are haemoglobin, haemostasis, urea and electrolytes, urinalysis (pH, ketones, protein, blood), random blood glucose, liver function tests, sickle cell screening, chest radiograph and ECG. For ASA 2 and 3, investigations include arterial blood gases, echocardiogram and lung function tests.

In general preoperative tests are not required in children undergoing minor or intermediate complexity procedures. Young healthy adults undergoing such procedures require limited preoperative investigation. Older patients and those with significant comorbidity undergoing major or major+ grade surgery require individualized preparation. Assistance on this issue is available in the form of evidence-based ‘look-up’ tables in the UK from the National Institute for Health and Clinical Excellence.

A respiratory opinion, pulmonary function tests and blood gas analyses are advisable in patients with respiratory disease that limits function and in patients undergoing thoracotomy. Forced expiratory volume in 1 second and forced vital capacity are good indices of obstructive and restrictive airways disease and can easily be measured. Similarly, a cardiology opinion and tests of cardiac function (24 hour ECG monitoring for arrhythmia, stress ECG, echocardiography, thallium scanning) are required in patients with intermediate predictors of risk for perioperative cardiac morbidity. The exact assessment for estimation of risk in the individual patient is left to the discretion of the cardiologist.

Risk assessment scoring systems

For the majority of patients, assessment of risk is based largely on clinical assessment. In essence, this is a judgement based on history and examination of the clinical, physiological and nutritional state of the patient. If carried out by an experienced clinician, this overall assessment may be as reliable as any complex scoring system. Usually, the clinical assessment is supplemented by taking into consideration the influence of individual factors (variables) that are known to have a documented adverse effect on outcome, e.g. old age, respiratory disease, cardiac disease and renal impairment. This type of additional assessment of risk is known as univariate, as the individual risk factors are considered one at a time. By contrast, multivariate (multifactorial) assessments provide a cumulative account (score) made up of the collective contributions of various data (clinical and laboratory), which reflect the overall risk and therefore the likely outcome. Some of these are generic, i.e. they can be used across a wide range of disease states, and others refer to specific disease processes and are relevant only in specific settings.

Generic scoring systems

American Society of Anesthesiologists grade

The term ASA refers to the scoring system described in the 1960s by the American Society of Anesthesiologists to assess fitness before surgery. It provides a measure of the preoperative physical status and is summarized in Table 3.1. Initially, the system was introduced to describe and select patients for clinical trials, but it has now been adopted for routine clinical use. It has the advantage of simplicity as it does not employ specific parameters. As such, grading of the patient relies on a certain degree of subjectivity. Furthermore, ASA grades may fail to adequately describe either patients with local, as opposed to systemic, disease (e.g. myocardial infarct) or those with multiple comorbidities of differing severity. This system is, however, used widely in surgical practice.

Table 3.1 Physical status scale of the American Society of Anesthesiologists (ASA)

Class	Physical status
1	A normally healthy individual: no organic, physiological, biochemical or psychiatric disturbance
2	A patient with mild to moderate systemic disease: this may or may not be related to the disorder requiring surgical treatment, e.g. diabetes mellitus, hypertension
3	A patient with severe systemic disease that is not incapacitating, e.g. heart disease with limited exercise tolerance, uncontrolled hypertension or diabetes
4	A patient with incapacitating systemic disease that is a constant threat to life with or without surgery, e.g. congestive cardiac failure, severe and persistent angina
5	A moribund patient who is not expected to live and where surgery is performed as a last resort, e.g. ruptured aortic aneurysm
E	A patient who requires an emergency operation

APACHE II

Assessment of outcome is an important consideration in patients requiring ICU treatment, for several reasons. First, intensive care is an extremely costly resource. The cost–benefit ratio cannot therefore be ignored and it is generally agreed that intensive care is inappropriate when the probability of survival is negligible. Within any society, the facilities for intensive care treatment are finite. If a selective policy is not adopted, the available resource becomes readily exhausted and salvageable patients who become acutely ill and require intensive care support may be denied this treatment. Thus, a policy of admitting inappropriate patients to the ICU, aside from being wasteful, is counterproductive.

There has been considerable clinical research into methods that afford reliable estimates for outcome in ICU patients. The method that has been most validated and is now widely accepted is the Acute Physiology and Chronic Health Evaluation (APACHE) system. It measures the severity of the acute disease by quantifying the degree of abnormality across multiple physiological variables. The APACHE system gives a score, which is the sum total of:

- acute physiological score (APS)
- age points
- chronic health points.

The maximum possible score with a revised APACHE system (APACHE II) is 71. In practice, no patient exceeds 55, and scores in excess of 35 are associated with a mortality exceeding 85% (Table 3.2).

Acute physiological score

The original APACHE I system comprised 34 potential physiological variables, making up the APS. As this was found to be impracticable, the infrequently measured and redundant variables have been dropped in the revised APACHE II, so that the APS is now made up from 12 variables (Table 3.2). The weighting system is based on a scale of 0 (normal range) to 4 (high or low abnormal). In practice, the APS is assessed on these variables over a 24 hour period, when the most deranged physiological value for each parameter for that day is used to calculate the score, e.g. the lowest blood pressure recording, lowest pH and highest respiratory rate.

Age points

Chronological age is an independent variable in its own right and, for this reason, points are assigned to the age in years as follows: 44 and below (0), 45–54 (2), 55–64 (3), 65–74 (5), 75 and over (6).

Chronic health points

As outcome is also adversely influenced by previous history of severe organ or systemic disorders and immunodeficiency states, points are allocated for these chronic disease problems. The latter disorders must have been present before the current severe illness in order to merit inclusion. The risk of death is higher when acutely ill patients with these previous chronic health problems require emergency surgical or non-operative treatment and when they undergo elective surgical operations and are electively admitted to the ICU directly from the recovery room. For this

Table 3.2 The Acute Physiology and Chronic Health Evaluation (APACHE) II severity of disease classification system

Physiological variable	**Score**									
	High abnormal range					**Low abnormal range**				
	+4	**+3**	**+2**	**+1**	**+0**	**+1**	**+2**	**+3**	**+4**	
Temperature; rectal (°C)	≥41	39–40.9		38.5–38.9	36–38.4	34–35.9	32–33.9	30–31.9	≤29.9	
Mean arterial pressure (mmHg)	≥160	130–159	110–129		70–109		50–69		≤49	
Heart rate (ventricular response)	≥180	140–179	110–139		70–109		55–69	40–54	≤39	
Respiratory rate (non-ventilated or ventilated)	≥50	35–49		25–34	12–24	10–11	6–9		≤5	
Oxygenation: A–aDO_2 or PaO_2 (mmHg)										
(a)	FiO_2 ≥0.5, record A–aDO_2	≥500	350–499	200–349		<200				
(b)	FiO_2 <0.5, record only PaO_2				PO_2 >70	PO_2 61–70		PO_2 55–60	PO_2 <55	
Arterial pH	≥7.7	7.6–7.69		7.5–7.59	7.33–7.49		7.25–7.32	7.15–7.24	<7.15	
Serum sodium (mmol/L)	≥180	160–179	155–159	150–154	130–149		120–129	111–19	≤110	
Serum potassium (mmol/L)	≥7	6–6.9		5.5–5.9	3.5–5.4	3–3.4	2.5–2.9		<2.5	
Serum creatinine (mg/100 mL) (double point scored for acute renal failure)	≥3.5	2–3.4	1.5–1.9		0.6–1.4		<0.6			
Haematocrit (%)	≥60		50–59.9	46–49.9	30–45.9		20–29.9		<20	
White blood count (total/mm^3) (in 1000s)	≥40		20–39.9	15–19.9	3–14.9		1–2.9		<1	
Glasgow Coma Scale (GCS) score										
Score = 15 minus actual GCS score										
(A) Total Acute Physiology Score (APS): sum of the 12 individual variable points	≥52	41–51.9		32–40.9	22–31.9		18–2.9	18–2.9	15–17.9	<15
Serum HCO_2 (venous; mmol/L) (not preferred, use if no ABG)										

(B) Age points: assign points to age as follows:

Age (years)	Points
≤44	0
45–54	2
55–64	3
65–74	5
≥75	6

(C) Chronic health points

If the patient has a history of severe organ system insufficiency or is immunocompromised assign points as follows:

(a) for non-operative or emergency postoperative patients: 5 points or

(b) for elective postoperative patients: 2 points

Definitions

Organ insufficiency or immunocompromised state must have been evident *prior* to this hospital admission and conform to the following criteria:

Liver	Biopsy-proven cirrhosis and documented portal hypertension: episodes of past upper GI bleeding attributed to portal hypertension; or prior episodes of hepatic failure/encephalopathy/coma
Cardiovascular	New York Heart Association class IV
Respiratory	Chronic restrictive, obstructive or vascular disease resulting in severe exercise restriction, i.e. unable to climb stairs or perform household duties; or documented chronic hypoxia, hypertension (>40 mmHg) or respirator dependency
Renal	Receiving chronic dialysis

Table 3.2 (Continued)

Immunocompromised	The patient has received therapy that suppresses resistance to infection, e.g. immunosuppression, chemotherapy, radiation; long-term or recent high-dose steroids, or has a disease that is sufficiently advanced to suppress resistance to infection, e.g. leukaemia, lymphoma, AIDS

APACHE II score

Sum of (A) + (B) + (C)

(A) APS points ....

(B) Age points ....

(C) Chronic health points ....

Total APACHE II..

ABG, arterial blood gas; GI, gastrointestinal.

reason, five points are awarded to non-operative or emergency postoperative patients as opposed to two points for elective postoperative patients with the same chronic health problems.

Mortality and APACHE II scores in patients treated in intensive care units

The relationship of the APACHE II system with mortality in patients admitted to ICUs is illustrated in Figure 3.1. There is a clear-cut inverse correlation between APACHE scores and survival; when scores exceed 29, the mortality increases sharply, and above 35 the chances of survival are so small as to be negligible. Within a given APACHE score range, the overall risk of hospital death varies in accordance with the underlying aetiology of the acute disease. Thus, in the score range 10–19, ICU patients with gastrointestinal bleeding or septic shock have a mortality of 27–30%, as distinct from 13% in patients with congestive cardiac failure.

POSSUM

While the APACHE II system is ideal for predicting mortality in the ICU, it is not so useful for the prediction of mortality and morbidity across the general surgical spectrum. The reasons for this are that it requires 24 hours of observation and the analysis of a large number of variables. As such, a simple scoring system for general surgical patients whose main use would be in surgical audit was developed. To start with, 62 factors were assessed by multivariate discriminate retrospective analysis. Of these, 35 covariates were then assessed prospectively to produce a system that has been termed Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM). This is divided into two sections: the physiological score, to be scored at the time of surgery, and the operative severity score. These are detailed in Tables 3.3 and 3.4.

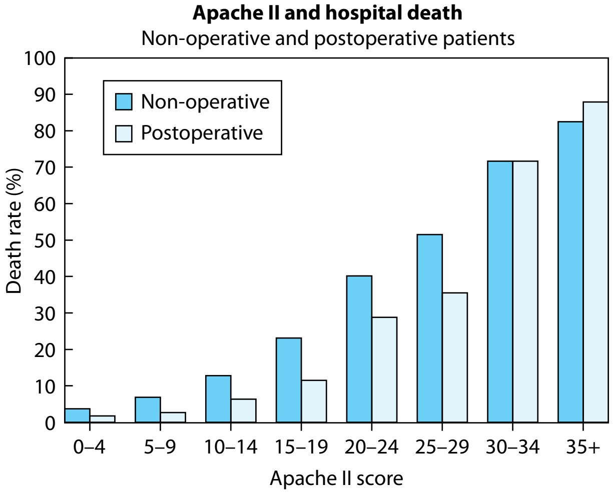


Figure 3.1 Relationship between Acute Physiology and Chronic Health Evaluation (APACHE) II scores and hospital mortality ($n = 5815$). (From Knaus *et al.*, *Crit Care Med* 1985;13:818–829.)

Thus, POSSUM requires scores from 12 physiological and six operative severity variables to estimate the risk of postoperative morbidity or mortality. For morbidity, the equation is:

$$\ln[R/(1-R)] = -5.91 + (0.16 \times \text{physiological score}) + (0.19 \times \text{operative severity score})$$

For mortality, the risk equation is:

$$\ln[R/(1-R)] = -7.04 + (0.13 \times \text{physiological score}) + (0.16 \times \text{operative severity score})$$

Prospective studies have shown POSSUM to be accurate in predicting both morbidity and mortality. Early criticism that POSSUM overpredicted mortality in low-risk groups have not been substantiated in later studies. These concerns did, however, lead to subsequent development of the Portsmouth, or P-POSSUM, predictor of mortality. Similar reports of overprediction of death in surgical subspecialties have also led to the development of individualized vascular (V-POSSUM) and colorectal (CR-POSSUM) scoring systems.

Disease-specific scoring systems

Over the years many scoring systems for different disease states or forms of trauma have been developed. In order to provide a representative example, the commonly used scoring systems for cirrhosis of the liver, pancreatitis, head injury, general trauma and tumour prognosis are described below.

Table 3.3 Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM): physiological score (to be scored at the time of surgery)

Parameter	Score			
	1	2	4	8
Age (years)	<60	61–70	>71	
Cardiac signs	No failure	Diuretic, digoxin, antianginal or hypertensive therapy	Peripheral oedema; warfarin therapy	Raised jugular venous pressure
Chest radiograph		Borderline cardiomegaly	Cardiomegaly	
Respiratory history	No dyspnoea	Dyspnoea on exertion	Limiting dyspnoea (one flight)	Dyspnoea at rest (rate >30/min)
Chest radiograph		Mild COAD	Moderate COAD	Fibrosis of consolidation
Blood pressure (systolic) (mmHg)	110–130	131–170	>171	–
		100–109	90–99	<89
Pulse (beats/min)	50–80	81–100	101–120	>121
		40–49		<39
Glasgow Coma Scale score	15	12–14	9–11	<8
Haemoglobin (g/100 mL)	3–16	11.5–12.9	16.1–17.0	10.0–11.4
		17.1–18.0	<9.9	>18.1
White cell count ($\times 10^{12}$/L)	4–10	10.1–20.0	>20.1	
		3.1–4.0	<30	
Urea (mmol/L)	<7.5	7.6–10.0	10.1–15.0	>15.1
Sodium (mmol/L)	>136	131–135	126–130	<125
Potassium (mmol/L)	3.5–5.0	3.2–3.4	2.9–3.1	<2.8
		5.1–5.3	5.4–5.9	>6.0
Electrocardiogram	Normal		Atrial fibrillation (rate 60–90)	Any other abnormal rhythm or >5 ectopics/min; Q waves or ST–T wave changes

COAD, chronic obstructive airways disease.

Table 3.4 Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM): operative severity score

Parameter	Score			
	1	2	4	8
Operative severity	Minor	Moderate	Major	Major+
Multiple procedures	1		2	>2
Total blood loss (mL)	<100	101–500	501–999	>1000
Peritoneal soiling or blood	None	Minor (serous fluid)	Local pus	Free bowel content, pus
Presence of malignancy	None	Primary only	Nodal metastases	Distant metastases
Mode of surgery	Elective		Emergency resuscitation of >2 hours possible	Emergency (surgery <2 hours needed)

Definitions of surgical procedures with regard to severity are guidelines; not all procedures are listed and the closest should be selected.

Cirrhosis of the liver

In cirrhosis of the liver, the prognosis is largely dependent on liver function. This is particularly important in assessing the risk of surgical procedures on patients with cirrhosis. Various grading systems have been devised. The most widely used is that of Child, modified by Pugh (Table 3.5). Grade A indicates a good prognosis, grade B a moderate prognosis and grade C a poor prognosis. An alternative classification is that of the Paul Brousse Hospital devised by Bismuth (Table 3.6).

Acute pancreatitis

In acute pancreatitis the diagnosis is confirmed by a rise in the serum amylase. Although amylase levels can be used for diagnosis they are not prognostic. It is also important in this context to estimate disease severity in order to rationalize treatment. Decisions relating to whether or not high-dependency or intensive care placement is indicated, use of antibiotics and whether early endoscopic retrograde cholangiopancreatography in patients with gallstone disease is indicated can be made upon the basis of whether patients are predicted to be at substantial risk of mortality or septic complications. For this reason, scoring systems have been developed using biochemical parameters. In 1992 the Atlanta classification was introduced. Although its limitations are widely reported, it permits classification of pancreatitis into mild and severe disease. Systems that predict for severity and the development of complications are the Ranson and Glasgow criteria. These scoring systems were developed using US and UK populations, respectively. They are summarized in Table 3.7. Ranson/Glasgow scores ≥3 define severe (necrotizing) pancreatitis and predict a protracted clinical course along with a higher mortality risk.

Scoring system for the assessment of head injury

One of the most frequent problems encountered by the clinician in the emergency department is to decide whether a patient is responding appropriately and is fully aware of their surroundings.

Table 3.5 Pugh modification of Child's classification system for assessing severity of hepatic dysfunction in cirrhosis

Parameter	Points awarded for abnormality		
	1	2	3
Encephalopathy	None	1–2	3–4
Ascites	Absent	Slight	Moderate
Albumin (g/L)	35	28–35	<28
Prothrombin time (seconds; prolonged)	14	4–6	6+
Bilirubin (mmol/L)	15–30	30–45	45+
Bilirubin patients in primary biliary cirrhosis	15–60	60–150	150+

Grade A, 5–6 points; grade B, 7–9 points; grade C, 10+ points.

Table 3.6 Paul Brousse Hospital classification of severity of chronic liver disease

Parameter	Number of criteria
Albuminaemia <3.0 g/100 mL	1
Hyperbilirubinaemia >30 mol/L	1
Encephalopathy	1
Clinical ascites	1
Coagulation factor II and V 40–60%	1
Coagulation factor II and V <40%	2

A, none of the criteria; B, one or two criteria; C, three or more criteria.

Table 3.7 Prognostic grading systems for acute pancreatitis

Ranson	Glasgow/Imrie
At admission/diagnosis	Within 48 hours
Age >55 years	Age >55 years
WBC >16 000/mm^3	WBC >15 000/mm^3
Blood glucose >10 mmol/L	Blood glucose >10 mmol/L (not diabetic)
AST >250 SF units/dL	Serum albumin <32 g/L
LDH >350 IU/L	Blood urea >16 mmol/L (no response to i.v. fluids)
Within 48 hours	LDH >600 IU/i
Haematocrit fall >10%	PaO$_2$ <60 mmHg
Blood urea nitrogen rise >5 mg/dL	AST/ALT >100 IU/L
Serum calcium <2.0 mmol/L	Serum calcium <2.0 mmol/L
PaO$_2$ <60 mmHg	
Base deficit >4 mEq/L	
Estimated fluid sequestration >61	

Three or more positive criteria in either system predicts a severe attack of acute pancreatitis.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; i.v., intravenous; LDH, lactate dehydrogenase; WBC, white blood cells.

Table 3.8 The Glasgow Coma Scale

Parameter	Score
Eye opening (E)	
Spontaneous	4
To speech	3
To pain	2
Nil	1
Best motor response (M)	
Obeys	6
Localizes	5
Withdraws (flexion)	4
Abnormal flexion	3
Extensor response	2
Nil	1
Verbal response (V)	
Orientated	5
Confused conversation	4
Inappropriate words	3
Incomprehensible sounds	2
Nil	1

Coma score: (responsiveness sum = 3–15 (E + M + V)).

In addition, difficulty in describing a patient's condition prohibits clear communication. The terms 'semiconscious', 'semicomatose' and 'stuporose' have different meanings to different people and so it is important to be as precise as possible regarding clinical findings. Fully conscious implies being fully aware of and interacting in an appropriate fashion to one's environment. If the patient is not fully conscious, the clinician must be able to describe the degree of consciousness.

The Glasgow Coma Scale (Table 3.8) uses (1) eye opening, (2) best motor response and (3) verbal response to define the level of consciousness. This has the advantage of providing an objective means of assessing the level of consciousness that is highly reproducible. It can therefore be used to monitor a patient's progress.

Injury severity score

In trauma patients the extent of injury and the probability of survival can be defined using the Injury Severity Scoring (ISS) system. The ISS system is based upon the Abbreviated Injury Score. In the ISS system, the body is categorized into six regions. Points are assigned to major organ systems depending on the type and severity of injury, i.e. minor = 1, moderate = 2, severe but not life threatening = 3, life threatening but survival probable = 4, and survival not probable = 5. Maximal injuries (currently untreatable cardiovascular, CNS or burn injuries) are allocated a score of 6. The three highest injury scores are then squared and summated. The ISS system scores range between 1 and 75. If any of the injuries score 6, the overall score defers to 75. Scores of around 25–40 are associated with a 50% mortality, depending on the patient's age.

Risk reduction and patient optimization

In recent years attention has focused on mechanisms that may reduce perioperative risk. The benefit of these strategies is likely to be greatest in high-risk groups.

β-blockade

β-adrenergic blockade in the weeks preceding elective surgery may confer postoperative outcome benefit in very high-risk patients. Influential trials have shown that pre-emptive β-blockade may reduce the risk of myocardial infarction and death in the context of vascular and non-cardiac surgery. These studies led to widespread use of this approach in the USA among high-risk patients. However, subsequent trials have suggested that the findings from initial studies should be interpreted with caution. In fact, the results from the Perioperative Ischemic Evaluation (POISE) trial showed that commencing β-blockers acutely on the day of surgery may in fact be harmful. National guidelines in the USA now recommend that β-blockade may be advantageous in patients at high risk of cardiac complications undergoing vascular and certain types of non-cardiac surgery.

Statins

A number of recent studies have demonstrated a possible mortality benefit to preoperative statin therapy in patients with ischaemic heart disease undergoing vascular surgery. One such retrospective observational study reported improved clinical outcome and resource utilization in patients receiving statin therapy and undergoing elective abdominal aortic aneurysm repair. Not all studies have, however, substantiated this cardiovascular effect.

Goal-directed fluid therapy

Previous studies have debated optimal perioperative fluid volume and type. Evidence in favour of both liberal and restrictive regimes utilizing either colloid or crystalloid fluid therapy are reported. Recent attention has, however, sought to improve understanding of the haemodynamic changes that occur during surgery.

An appropriate physiological response to the stress of surgical intervention is to increase cardiac output. Shoemaker and colleagues demonstrated that patients incurring a significant oxygen debt during surgery are at high risk of postoperative complications. Patients who sustain a large and persistent oxygen debt are at greatest risk of mortality. The aim of goal-directed therapy is to maintain adequate tissue perfusion. Correction of physiological variables to within desirable target ranges is feasible with judicious use of fluids, oxygen and inotropic agents [dobutamine, dopexamine and norepinephrine (noradrenaline)]. Perioperative monitoring modalities offer real-time information on heart rate, urine output, oxygen saturation, stroke volume, arterial blood pressure, central venous oxygen saturation and cardiac output, thereby permitting manipulation. Oesophageal Doppler is one such method of measuring intraoperative cardiac output during abdominal surgery that has gained significant popularity. Goal-directed therapy has been associated with reduced mortality and shorter length of stay among high-risk groups undergoing major surgery. It has been advocated that high-risk patients (i.e. those at considerable risk of postoperative mortality) benefit most from goal-directed therapy. Therefore, some degree of patient selection in this context seems appropriate.

Perioperative nutrition

In recent years perioperative Enhanced Recovery of patients After Surgery (ERAS) has been incorporated into mainstream clinical practice. Nutritional aspects of care are central to such multimodal rehabilitation protocols. The severely malnourished patient (weight loss of ≥10–15% body weight, BMI <18.5, albumin <30 g/dL) should be offered nutritional support for at least 10–14 days even if this results in delayed surgery. Even patients who are not malnourished but are anticipated to not be able to feed normally for 1 week should be considered for nutritional support. The key principles of perioperative nutrition management include: minimization of the catabolic response to surgery, avoidance of long periods of starvation with reinstitution of feeding as soon as possible after surgery. It is now common practice to permit patients to drink clear fluids until 2 hours before surgery and the avoidance of solids is limited to 6 hours preoperatively. In addition, patients undergoing major surgery are offered carbohydrate loading drinks in the hours preceding the intervention. After surgery, patients should be encouraged to resume normal food intake or enteral feeding as soon as possible. The majority of patients can be fed early by mouth even after major gastrointestinal procedures. Certain groups, however, including those undergoing head and neck surgery, severe trauma patients, patients who are malnourished at the time of surgery and those who will not manage oral intake for 7–10 days, will require enteral tube or parenteral nutrition. The enteral route is preferred. Enteral feeding formulas comprising standard protein regimes are used commonly. Perioperative supplementation with immune-modulating substrates such as arginine, ω-3 fatty acids and nucleotides is beneficial in certain circumstances (severe trauma and surgery for upper gastrointestinal malignancy).

GUIDE TO FURTHER READING

Bradley EL III. A clinically based classification system for acute pancreatitis. Summary of the International Symposium on Acute Pancreatitis, Atlanta, September 11–13, 1992. *Arch Surg* 1993;**128**:586–590.

Carson JL, Spence RK, Poses RM, Bonavita G. Severity of anaemia and operative mortality and morbidity. *Lancet* 1988;**331**:727–729.

Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991;**78**:356–360.

Douketis JD, Berger PB, Dunn AS, *et al.* The perioperative management of antithrombotic therapy: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). *Chest* 2008;**133**(6 suppl.):299S–339S.

Fleisher LA, Beckman JA, Brown KA, *et al.* ACC/AHA 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery): developed in collaboration with the American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Rhythm Society, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society for Vascular Medicine and Biology, and Society for Vascular Surgery. *Circulation*

2007;**116**:e418–499. Erratum in: *Circulation* 2008;**118**: e143–144; *Circulation* 2008;**117**:e154.

Fleischmann KE, Beckman JA, Buller CE, *et al*. ACCF/AHA 2009 focused update on perioperative beta blockade. a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation* 2009;**120**:2123–2151.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818–829.

Lee TH, Marcantonio ER, Mangione CM, *et al*. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999;**100**:1043–1049.

Mangano DT. Perioperative cardiac morbidity. *Anesthesiology* 1990;**72**:153–184.

National Institute for Health and Clinical Excellence. *Preoperative Tests*. Clinical Guideline no. 3. London, UK: NICE, 2003. Available from: http://www.nice.org.uk/nicemedia/live/10920/29090/29090.pdf

Shoemaker WC, Appel PL, Kram HB. Tissue oxygen debt as a determinant of lethal and nonlethal postoperative organ failure. *Crit Care Med* 1988;**16**:1117–1120.

Tekkis PP, Prytherch DR, Kocher HM, *et al*. Development of a dedicated risk-adjustment scoring system for colorectal surgery (colorectal POSSUM). *Br J Surg* 2004;**91**:1174–1182.

Weimann A, Braga M, Harsanyi L, *et al*. ESPEN guidelines on enteral nutrition: surgery including organ transplantation. *Clin Nutr* 2006;**25**:224–244.

Whiteley MS, Prytherch DR, Higgins B, *et al*. An evaluation of the POSSUM surgical scoring system. *Br J Surg* 1996;**83**:812–815.

CHAPTER 4

Care of the patient in the perioperative period

HELGI JOHANNSSON AND VAFA MANSOUBI

The preoperative preparation of a patient for surgery

The perioperative period is a time of increased risk for every patient undergoing an operation. This risk is caused by the effects of anaesthesia, the operation itself and its complications, and the stress response to surgery affecting the patient's immune and coagulation status. Balancing the risk of the surgery and the benefit to the patient can be a difficult task that will involve the surgeon, anaesthetist and often the patient's physician, as well as the patient themselves.

Factors affecting the risk of surgery

These include:

- the nature of the surgery
- pre-existing medical conditions
- the patient's physical fitness
- whether the surgery is planned or an emergency
- presence of a coagulation disorder
- obesity
- age
- smoking.

Preassessment of patients presenting for surgery

The Association of Anaesthetists of Great Britain and Ireland (AAGBI) recommends that anaesthetists should assume a central role in the preoperative preparation of patients for operations. Preassessment clinics have been set up in most hospitals in the UK using anaesthetists and specifically trained nurses to screen for conditions that may increase the patient's perioperative risk. Their aim is not only to pick up these conditions but to give patients information about their operation and anaesthetic, as well as discontinue relevant drugs and advise patients on starvation times.

Cardiovascular disease

The perioperative period places unique stresses on the cardiovascular system. Anaesthesia itself causes great changes in blood pressure, systemic vascular resistance and myocardial contractility. As a consequence, blood pressure routinely falls on induction of anaesthesia owing to vasodilation and suppression of myocardial contraction. Intubation and the start of the operation often increase the heart rate and blood pressure, putting further stress on the system. Cardiovascular disease is common in the general population. Pre-existing cardiac disease increases the risk of perioperative myocardial infarction, heart failure and even complete loss of cardiac output.

Cardiovascular issues affecting surgery

The cardiovascular issues affecting surgery include:

- ischaemic heart disease
- previous angioplasty
- anticoagulation and antiplatelet therapy
- hypertension
- heart failure
- atrial fibrillation and other arrhythmias
- valvular heart disease.

Table 4.1 Tests used in diagnosing and assessing the severity of heart disease

Test	Use
ECG	Previous MI, hypertrophy, conduction defects, arrhythmias
Echocardiogram	Heart failure, hypertrophy, valvular heart disease
Stress echocardiogram	Above plus inducible cardiac ischaemia
24 hour ECG monitoring	Paroxysmal arrhythmias
Exercise ECG	Inducible cardiac ischaemia
Angiogram	Coronary artery disease and intervention
Cardiopulmonary exercise testing	Global fitness and oxygen usage. Heart failure, respiratory failure. Often difficult to interpret

MI, myocardial infarction.

Ischaemic heart disease

Plaque in the coronary arteries causes a restriction in oxygen supply to the myocardium, particularly at times of high demand. The heart muscle is perfused in diastole owing to the high pressures in the myocardium during systole, and oxygen supply is therefore reduced if the heart rate increases significantly. High heart rates and shear forces on the arteries brought about by increased contractility may cause a coronary plaque to rupture and consequently myocardial infarction. In addition, the hypercoagulable state in the perioperative period can increase the risk of infarction.

Risk factors for ischaemic heart disease

These are:

- smoking
- diabetes
- family history
- hypercholesterolaemia
- ethnic groups (e.g. Indian subcontinent)
- male sex.

Detecting and optimizing ischaemic heart disease

The history and an appreciation of the patient's risk factors are vital in the detection of coronary arterial disease. The patient may dismiss the symptoms as trivial and concentrate more on the reason for surgery. Angina increases the patient's risk of myocardial infarction perioperatively, and the presence of heart failure increases mortality.

The tests shown in Table 4.1 are of use in diagnosing and assessing the severity of heart disease. Their preoperative use is controversial. However, cardiopulmonary exercise testing has been used in evaluating perioperative risk, and in particular a low anaerobic threshold has been found to be a poor prognostic indicator in major surgery. Interpretation of cardiopulmonary exercise testing is often difficult and poor patient effort may produce misleading results.

Arrhythmias

Optimization of angina

This should be done in close liaison with the patient's cardiologist or general practitioner. If the patient's angina is stable and the patient is on appropriate medical treatment, surgery is presently not in itself an indication for revascularization. The use of coronary stents prior to non-cardiac surgery necessitates a period of antiplatelet therapy with aspirin and clopidogrel, which in itself may increase the morbidity and mortality of surgery.

Valvular heart disease

Valvular heart disease can influence greatly the fitness of a patient for surgery. Of particular note are stenotic valves, which will greatly limit the ability of the body to increase its cardiac output. The most common of these is aortic stenosis. As mentioned before, anaesthesia causes marked vasodilation, leading to a reduction in peripheral vascular resistance. Normally, this would be compensated for by an increase in cardiac output, but with a stenotic valve in place this is not possible. Therefore, there will be an exaggerated and sometimes catastrophic reduction in the blood pressure.

Aortic stenosis

Aortic stenosis is:

- often asymptomatic, but may present with syncope or angina
- identified by an ejection systolic murmur, associated slow rising pulse and heaving apex beat; there is narrow pulse pressure
- characterized by left ventricular hypertrophy on ECG
- diagnosed by echocardiogram.

Mitral stenosis

Mitral stenosis:

- presents with atrial fibrillation (AF), syncope, general malaise
- has a subtle diastolic murmur, which is often missed
- shows a large p wave (p mitrale) on ECG – if the patient is in sinus rhythm
- diagnosed by echocardiogram.

Arrhythmias

Atrial fibrillation

The most common arrhythmia in the general surgical population, AF is present in approximately 1% of the UK population, increasing to 10% in those over 75 years old. The atria contribute approximately 15% of the total cardiac output, and atrial fibrillation increases the likelihood of blood stagnating in the atrium, clot formation and distal embolization. A patient in AF is therefore at increased risk perioperatively of both low cardiac output states and distal ischaemia owing to embolism. Paroxysmal AF may be difficult to diagnose as ECG evidence may not be available, but has a similar risk of embolism. Most patients who have been diagnosed with AF are anticoagulated (http://www.bcshguidelines.com/documents/warfarin_4th_ed.pdf) and so present an additional challenge in the perioperative period.

Risk factors for perioperative atrial fibrillation

These include:

- major surgery
- electrolyte abnormalities:
  - hypokalaemia
  - hypomagnesaemia
- infection (especially pneumonia)
- central venous catheters (if inserted too far)
- advanced age
- previous atrial fibrillation episode

- thyrotoxicosis
- Wolff–Parkinson–White and other disorders of conduction.

Prevention and treatment of perioperative atrial fibrillation
This includes:

- keeping the serum potassium concentration above 4.5 mmol/L
- treating hypomagnesaemia early
- ensuring correct placement of central venous catheters
- treating sepsis early
- considering antiarrhythmics in liaison with a physician or anaesthetist
- considering DC cardioversion.

Narrow complex tachycardia
Syndromes such as Wolff–Parkinson–White may lead to perioperative re-entry tachycardia (supraventricular tachycardia). A history of fast palpitations and previous cardioversions may point to this diagnosis. A cardiology review and continuing antiarrhythmics is important in these patients.

Heart failure

Heart failure is defined as the inability of the heart to deliver sufficient oxygen to the body to satisfy its metabolic demands. At rest, it may be asymptomatic, or manifest itself only in non-specific symptoms such as fatigue. Heart failure poses a significant risk for major and minor surgery as the body is not able to respond to an increase in the metabolic demands and haemodynamic changes that occur during surgery and anaesthesia.

Diagnosis
History and examination are vital in the diagnosis of latent heart failure, and the severity of heart failure is graded according to the New York Heart Association classification (Figure 4.1). Echocardiography and cardiopulmonary exercise testing are useful in the investigation and risk classification of heart failure.

Hypertension

Hypertension is persistent raised blood pressure above 140/90 mmHg. It is extremely common in the surgical population. Hypertension is a risk factor for ischaemic heart disease, cerebrovascular disease and renal disease.

'White coat hypertension' is the presence of elevated blood pressure at clinic/hospital visits only, with normal blood pressure when not agitated. It may be necessary to perform ambulatory blood pressure monitoring to distinguish this from genuine hypertension.

Mildly elevated blood pressures have not been shown to present a significantly increased risk for surgery in trials.

Patients with severe hypertension should have their elective surgery deferred for treatment of their blood pressure. First-line therapy of thiazide diuretics may not be appropriate for this as it frequently takes up to 6 weeks to see an improvement in the blood pressure.

Patients on antihypertensive therapy should continue this in the perioperative period – particularly beta-blockade. Being 'nil-by-mouth' does not preclude patients taking their tablets with a little water. Angiotensin-converting enzyme blockers may increase haemodynamic instability during surgery, and some anaesthetists recommend withholding these on the day of surgery.

Anticoagulation in the perioperative period

Anticoagulant and antiplatelet therapy is commonly encountered in patients undergoing surgery and presents a significant challenge to the surgeon and anaesthetist. Careful balancing of the risk of pausing anticoagulant or antiplatelet therapy with the risk of bleeding must be undertaken and a decision made whether to start the patient on short-term anticoagulation, or simply to stop long-term treatment and restart once the likelihood of bleeding is low.

Common indications for anticoagulation

These include:

- atrial fibrillation (permanent and paroxysmal)
- valvular heart disease
- previous venous thromboembolism
- thrombophilia
- bypass grafts and other foreign materials.

NYHA Class	Symptoms
I	Cardiac disease, but no symptoms and no limitation in ordinary physical activity, e.g. shortness of breath when walking, climing stairs etc.
II	Mild symptoms (mild shortness of breath and/or angina) and sliht limitation during ordinary activity.
III	Marked limitation in activity due to symptoms, even during less-than-ordinary activity, e.g. walking short distances (20-100m). Comfortable only at rest.
IV	Severe limitations. Experiences symptoms even while *at rest*. Mostly bedbound patients.

Figure 4.1 New York Heart Association classification of heart failure.

Common indications for antiplatelet therapy

These include:

- ischaemic heart disease
- previous myocardial infarction
- angioplasty and coronary stents (particularly drug-eluting stents).

Management of warfarin

Warfarin treatment is common in the surgical population. It acts by inhibition of synthesis of vitamin K-dependent clotting factors in the liver. It has a slow onset of action and requires a loading dose. It takes around 3–4 days for a warfarinized patient's clotting to return to normal after it has been stopped. The management of warfarinization perioperatively should be altered in relation to the risk of thromboembolism during that period. Other forms of anticoagulation (such as low-molecular-weight heparin) may have greater risks than warfarin, and it may not be in the patient's best interest to continue full anticoagulation before surgery. This is a guide to management:

- Low risk of thromboembolism (e.g. lone atrial fibrillation): simply withhold warfarin 4 days before operation (or according to local protocol) and restart once the risk of bleeding is low postoperatively.
- Intermediate risk (e.g. recurrent thromboembolism): stop warfarin as above, but cover with a treatment dose of low-molecular-weight heparin starting the day after warfarin has been stopped, the last dose being given on the morning of the day before surgery.
- High risk (e.g. prosthetic heart valves): as above but admit the patient the day before surgery and commence an unfractionated heparin infusion. Stop 2 hours preoperatively, measure the activated partial thromboplastin time, and restart as soon as the risk of bleeding is low.

It should be stressed that oral non-warfarin anticoagulants such as dabigatran are becoming more common, and may supersede injectable heparins in time (see Chapter 7).

Diabetes mellitus

The perioperative care of patients with diabetes mellitus causes much confusion and often leads to worsening of diabetic control at the time of surgery. In addition, the stress response to surgery and many of the drugs used during surgery and anaesthesia increase insulin resistance, worsening diabetic control.

Patients with type 1 diabetes often present to surgery for treatment of complications of their disease. These complications are more likely when control has been poor or the condition has been present for a long time.

Complications of diabetes mellitus

These are:

- ischaemic heart disease
- hypertension
- renal disease
- retinopathy
- cerebrovascular disease
- peripheral vascular disease
- peripheral and autonomic neuropathy
- hyper- and hypoglycaemia
- gastric stasis.

Principles of management of the diabetic patient in the perioperative period

These are as follows:

- minimize starvation time
- ensure good diabetic control before elective surgery [haemoglobin A_{1C} (HbA_{1C}) between 6% and 8%]
- ideally, the patient should be first on the list
- keep the patient on his or her normal regimen for as long as possible
- those with type 1 diabetes should have an insulin infusion for long operations.

CASE STUDY 1

A 46-year-old woman with type 1 diabetes is scheduled for a laparoscopic cholecystectomy. She is on a long-acting insulin, which she takes in the evening, and injects herself with short-acting insulin before every meal, monitoring her glucose after her meal, supplementing insulin when necessary. Her control is excellent and she is very familiar with her insulin regimen.

Traditionally, she would be admitted the night before and started on a sliding scale of insulin. However, in this case she is admitted on the day, advised to take her long-acting insulin as normal and can drink clear fluids until 06.30 in the morning. She is first on the operating list, her blood glucose is monitored intraoperatively and she is allowed to eat and drink afterwards, having her lunch on the ward at midday with her normal short-acting insulin. There is minimal disruption to her insulin regimen and she can be discharged once the normal criteria have been met.

CASE STUDY 2

A 58-year-old man is scheduled for a right hemicolectomy. He has type 1 diabetes with nephropathy and previous retinal surgery. He is on twice-daily injections of a long/short-acting insulin mix, and is envisaged to have a period of not eating for around 2 days after the operation. He is admitted the night before surgery and started on an insulin infusion sliding scale, and intravenous fluids to avoid dehydration. He could be admitted on the morning of surgery if he omits his morning insulin (having taken the previous night's dose), and started on insulin after induction of anaesthesia. Care must be taken to account for his normal insulin dose in the regimen he gets written up for rather than just to opt for a standard sliding scale if he is on an unusually high or low dose of the mixed insulin. Consultation with a diabetologist will be useful in this instance.

CASE STUDY 3

A 65-year-old man presents for a hernia repair. He has type 2 diabetes treated with metformin and gliclazide. He is obese with a body mass index of 33, and smokes occasionally. His creatinine is 145 mmol/L and he had a coronary stent inserted 2 years ago. He is presently on aspirin.

During his initial appointment he is encouraged to prepare for his operation and minimize his anaesthetic risk. This includes referral to a smoking cessation programme and advice on weight reduction and exercise. After a risk–benefit discussion between the surgeon and cardiologist the decision is made to continue aspirin perioperatively, and his oral hypoglycaemics are omitted on the day of operation. He is scheduled first on the operating list.

Because of his high creatinine level he does not have a non-steroidal anti-inflammatory drug prescribed, and he has his gliclazide with lunch. The metformin is withheld until the following morning because of the small risk of lactic acidosis.

Type 1 diabetes mellitus

Type 1 diabetes mellitus is an autoimmune-mediated disease characterized by the destruction of the cells of the pancreas, leading to the inability to produce endogenous insulin. The consequences of insulin deprivation are severe, starting with a metabolic ketoacidosis and osmotic diuresis.

Preoperative management: principles

Withholding insulin from patients with type 1 diabetes leads to acidosis and may prove fatal. It is therefore vital that in the perioperative period insulin administration is not stopped for any significant length of time.

Type 2 diabetes

Unlike type 1 diabetes, in which there is a lack of insulin, type 2 diabetes is characterized by resistance to insulin. Therefore, a lack of insulin does not lead to ketoacidosis, and the absence of insulin during periods of 'nil by mouth' are not as dangerous as in type 1 diabetes.

The key points are:

- optimize control of diabetes and aim for HbA_{1c} of 8% or below
- screen for end-organ damage such as renal impairment and ischaemic heart disease
- advise lifestyle modification (smoking cessation, exercise and weight loss)
- minimize starvation time
- withhold oral hypoglycaemic agents on the day of operation.

Respiratory disease

Assessment of respiratory disease

History and examination still form the mainstay of respiratory assessment, with exercise tolerance providing vital and valuable information on the patient's physical reserve. The ability to climb one flight of stairs without stopping is used by many anaesthetists to predict a trouble-free anaesthetic, but difficulties may arise when there are other disabilities making movement difficult, principally arthritis. Peak flow and pulmonary function tests provide further information but are again difficult to interpret when the patient has poor technique or limited understanding on how to perform these tests:

- peak flow: useful to screen for bronchospasm and reduced respiratory muscle power
- pulse oximetry: detects hypoxia
- spirometry: distinguishes between restrictive and obstructive disease
- full pulmonary function tests: also detect interstitial lung disease if transfer factor is included
- chest radiograph: of limited use as a preoperative screening test.

Chronic obstructive pulmonary disease

Chronic obstructive pulmonary disease (COPD) is the presence of chronic airflow limitation, excessive secretions and destruction of lung tissue, usually as a result of smoking. Patients with COPD have around three times higher risk of pulmonary complications after surgery. This rises to 4.7 times for thoracic and abdominal surgery. The decision on proceeding to surgery should be made by the consultant anaesthetist and surgeon looking after the patient, taking into account the severity of the disease and the risks and benefits of the operation (National Institute for Health and Clinincal Excellence 2010).

Asthma

Asthma is common in the surgical population, and characterized by reversible bronchoconstriction caused by smooth muscle activation in the bronchi. The spectrum of disease varies from very mild exercise-induced asthma with no impact on the day-to-day activity of the individual, to constantly life-threatening asthma with frequent admissions to hospital and intensive care units. The mild forms of the disease have little impact on the patient's perioperative risk, whereas the severe form increases the risk of operation and anaesthetic considerably.

Features suggesting severe asthma

These include:

- exercise tolerance limited by shortness of breath
- previous hospital admissions with asthma exacerbations
- high doses of inhaled steroids
- frequent courses of oral steroids
- second- and third-line asthma drugs
- fluctuations in peak flow.

Preparation for surgery: asthma and chronic obstructive pulmonary disease

Preparations include:

- history and examination
- considering pulmonary rehabilitation and optimization by a respiratory physician

- smoking cessation
- evaluation of the severity of COPD (spirometry, pulmonary function tests, exercise tolerance test)
- careful evaluation of the risks and benefits of surgery in the case of severe disease
- continuing respiratory medication perioperatively
- considering a nebulized bronchodilator immediately prior to surgery.

Obstructive sleep apnoea

During sleep, a person's airway muscle tone decreases and obstruction can develop (e.g. snoring). Obstructive sleep apnoea is a condition in which there are frequent pauses of 20–40 seconds during sleep when the patient does not take a breath. Hypoxia ensues and carbon dioxide builds up in the body. Symptoms may be subtle and commonly include daytime somnolence, tiredness and restless sleep. In severe cases, pulmonary hypertension and cor pulmonale may develop, leading to right heart failure and peripheral oedema.

The prevalence of obstructive sleep apnoea is increasing and is often undiagnosed in a patient attending for surgery.

Risk factors for obstructive sleep apnoea

These include:

- obesity
- anatomical features causing narrowing of the airway
- craniofacial syndromes
- weak airway muscle tone
- previous brain injury.

Obstructive sleep apnoea presents a particular challenge during the perioperative period. The anatomical features that lead to obstructive sleep apnoea can frequently lead to difficulties in airway management. In addition, during airway management, any difficulties may lead to trauma and swelling of an already narrowed airway. This in turn can lead to difficulties in breathing in the immediate postoperative period when the patient may already be drowsy from anaesthetic agents and analgesics. These patients are at higher risk postoperatively from upper airway obstruction and hypoventilation.

Gastrointestinal considerations

Anaesthesia obtunds the reflexes protecting the airway from foreign materials, including stomach contents. Anaesthesia reduces the tone of the gastro-oesophageal sphincter, and many anaesthetic agents cause nausea and vomiting. Stomach contents are very acidic and cause a chemical pneumonitis when aspirated into the respiratory tract.

Factors delaying gastric emptying

These are:

- large solid fatty meal
- distress and pain (particularly in children)
- labour
- position (supine)
- obstruction (e.g. gastric outflow obstruction)
- abdominal sepsis (e.g. appendicitis)
- drugs: opioids, anticholinergics, dopamine
- autonomic disorders (e.g. diabetes-induced gastroparesis).

Preventing aspiration of gastric contents

Strategies employed by anaesthetists to prevent aspiration pneumonia involve identifying patients at risk, premedication, modification of anaesthetic techniques and postoperative care aimed at minimizing drowsiness, nausea and vomiting (see Rapid sequence induction).

Central nervous system disorders

Cerebrovascular disease

Patients with previous strokes frequently present for anaesthesia for unrelated conditions. The perioperative period presents a high risk of further strokes and it is important to weigh carefully the benefit of the operation with the risk to the patient. This should be done during a careful preassessment and discussion between the surgeon and anaesthetist, with the patient and relatives closely involved.

Perioperative strategies for prevention of further stroke

These measures are:

- when possible continue antiplatelet therapy
- reinstate antiplatelet therapy (when stopped) as soon as is deemed safe
- ensure antihypertensives are continued perioperatively
- avoid hypotension and keep the patient's blood pressure within 20% of their preoperative blood pressure.

Disorders causing muscle weakness

Disorders such as motor neurone disease and myasthenia gravis lead to a reduction in skeletal muscle power. This may have severe implications for recovery from anaesthesia. The presence of respiratory muscle weakness increases the risk of postoperative hypoventilation; pneumonia, weak cough and weak upper airway reflexes increase the risk of aspiration of gastric contents:

- continue therapeutic medication (e.g. pyridostigmine) on the day of operation to maximize power postoperatively
- minimize medication that increases drowsiness, and use regional techniques when possible
- ensure there is minimal residual neuromuscular blockade remaining postoperatively
- the patient may require a high-dependency or intensive care unit bed postoperatively to monitor for hypoventilation and signs of sepsis
- the decision whether to operate or not should be taken by the surgeon and anaesthetist in discussion with the patient and should take into account the risks and the benefits of the surgery.

Intraoperative care

Principles of anaesthesia

Anaesthesia is a pharmacologically induced and reversible lack of awareness. This can be total lack of awareness, as in general anaesthesia, or partial, as in sedation.

General anaesthesia

This includes the triad of hypnosis, analgesia and muscle relaxation that is started in the period of *induction of anaesthesia* and further continued during *maintenance* until the end of the procedure, when *emergence* from anaesthesia begins.

Induction of anaesthesia

This is usually achieved by the use of *intravenous* anaesthetic agents, but the *inhalational* route using volatile anaesthetics may also be used.

Intravenous agents include propofol, thiopentone, etomidate, ketamine:

- advantages: rapid onset, titratable dose
- disadvantages: hypotension and apnoea common (except ketamine).

For inhalational induction, volatile agents are used, including sevoflurane, halothane and nitrous oxide:

- advantages: no need for initial cannulation (paediatrics), respiration maintained
- disadvantages: slower onset (especially in adults), unpleasant and irritable.

Rapid sequence induction

When a patient is judged to be at risk of aspiration of gastric contents a rapid sequence induction is used. The principles of the rapid sequence induction are to minimize the time between the loss of laryngeal reflexes and the securing of the airway. It is used to achieve optimal intubating conditions in around 30 seconds.

It involves:

- preoxygenation: 3–5 minutes, leading to replacement of nitrogen for oxygen in the functional residual capacity and increase oxygen stores to maximize the time available before the patient desaturates
- cricoid pressure: aims to prevent passive regurgitation of gastric contents
- availability of suction and tilting trolley: to minimize aspiration if regurgitation occurs
- fast administration of thiopentone and suxamethonium or rocuronium: fast induction and paralysis
- avoidance of bag–mask ventilation, which may lead to air entering the stomach and increasing gastric pressure.

Maintenance

Both the intravenous and inhalational route may be used.

Inhalational routes use volatile agents (sevoflurane, desflurane, isoflurane) in combination with carrier gases (oxygen, air, nitrous oxide):

- advantages: monitoring of end-tidal levels and so capillary (blood) levels; familiarity, convenience, minimal additional equipment needed
- disadvantages: emetogenic, greenhouse pollutant.

Intravenous routes can be infusion only (total intravenous anaesthesia) or boluses. Propofol is the only anaesthetic agent currently used for maintenance:

- advantages: reduced postoperative nausea and vomiting
- disadvantages: dedicated intravenous line, so it is not possible to monitor blood levels (risk of awareness).

Emergence

This is simply achieved by stopping the anaesthetic agent and administering high-flow oxygen. The inhalational agents are then expired, leading to eventual drop in blood concentration levels and subsequent emergence. Propofol has a very large volume of distribution and on cessation of the infusion will redistribute and lead to emergence.

Muscle relaxation

Anaesthetic agents provide a certain degree of muscle relaxation, but paralysis is achieved by giving neuromuscular blockers. This is to aid intubation and provide paralysis for easier surgical access (e.g. relaxation of abdominal muscles for laparotomy and laparoscopy).

They are classified into:

- Depolarizing neuromuscular blockers: suxamethonium:
  - advantages: fast paralysis (intubation in 30 seconds), short acting
  - disadvantages: numerous side effects, including myalgia, hyperkalaemia, bradycardia, increased intracranial and intraocular pressure.
- Non-depolarizing neuromuscular blockers: atracurium, vecuronium, rocuronium
  - advantages: safer side effect profile
  - disadvantages: long onset time, may accumulate with renal and hepatic failure.

Other agents

Analgesics (opioids, α_2-agonists) and other hypnotics (benzodiazepines, α_2-agonists) are added to reduce the anaesthetic agent requirements as well as provide analgesia.

Ketamine

This drug is gaining in popularity, especially in emergency and field anaesthesia. It acts as an *N*-methyl-ᴅ-aspartic acid antagonist and can cause profound analgesia, anaesthesia and amnesia. Its routine use is limited because of its potential side effects, including emergence delirium, dysphoria and hallucinations, tachycardia, hypertension and increased intracranial pressure.

Airway

During anaesthesia it is important to maintain an open airway owing to loss of normal tone causing the tongue and soft tissues to obstruct the airway.

Different types of airway include:

- Oropharyngeal/nasopharyngeal: least likely to stimulate airway reflexes but more prone to failure and stomach inflation with ventilation.
- Laryngeal mask: easy to insert requiring little training.
- Endotracheal tube (oral/nasal): definitive airway allowing efficient ventilation and protecting against airway soiling.

- Tracheostomy: a hole made in the trachea allowing bypass of the upper airway altogether. This is frequently used as a part of a weaning protocol on the intensive care unit. Also it is rarely performed as an emergency for a failed attempt at intubation (in this case usually through the cricothyroid membrane).

Patient may ventilate *spontaneously* but *artificial* ventilation is carried out during periods of apnoea or if paralysis is used.

The methods of ventilation include:

- bag and mask
- hand ventilation using anaesthetic circuits
- mechanical ventilation.

Monitoring during anaesthesia

There are standards of basic monitoring that must be carried out during the conduct of all anaesthetics. These have been recommended by the AAGBI and are strictly adhered to by all anaesthetists.

These apply to all procedures including local/regional and sedative techniques and should start before the induction of anaesthesia and continue until the patient has fully recovered.

The AAGBI guidelines on minimal standard monitoring include:

- the presence of anaesthetists throughout the procedure
- pulse oximetry
- non-invasive blood pressure (NIBP) monitoring
- ECG
- capnography and monitoring of gases: O_2, CO_2, vapours
- airway pressures.

With the availability of the following:

- means of measuring temperature
- nerve stimulator.

During the recovery phase, pulse oximetry and NIBP must be present and all others must be available.

The aims of standard monitoring during anaesthesia include:

- monitoring of vital signs to ensure adequate tissue perfusion and oxygenation
- monitoring depth of anaesthesia
- early warning of adverse events and errors.

Additional monitoring

Other forms of monitoring are also used in individual patients and during certain procedures. These include:

- urine output: indicator of cardiac output, volume status, kidney function
- invasive blood pressure monitoring: allow beat-to-beat assessment of blood pressure and sampling line for arterial blood gas measurements
- central venous monitoring: indication of patient volume status, effect of fluid challenge and venous blood sampling
- other modes of cardiac output measurement: Doppler probe, pulmonary artery flotation catheter, PICCO (pulse contour analysis) (see Chapter 6)
- depth of anaesthesia: electroencephalogram, modified EEG, bispectral analysis, evoked potentials.

World Health Organization surgical safety checklist

This is a worldwide initiative to reduce the number of surgical errors and complications. It has been endorsed by the Department of Health, National Patient Safety Agency and all the Royal Colleges and, after an initial trial, has now been implemented nationwide.

The WHO's 10 essential objectives include:

1. The team will operate on the correct patient at the correct site.
2. The team will use methods known to prevent harm from anaesthesia, while protecting the patient from pain.
3. The team will recognize and effectively prepare for life-threatening loss of airway or respiratory function.
4. The team will recognize and effectively prepare for risk of high blood loss.
5. The team will avoid inducing any allergic or adverse drug reaction known to be of significant risk.
6. The team will consistently use methods known to minimize risk of surgical site infection.
7. The team will prevent inadvertent retention of instruments or swabs in surgical wounds.
8. The team will secure and accurately identify all surgical specimens.
9. The team will effectively communicate and exchange critical patient information for the safe conduct of the operation.
10. Hospitals and public health systems will establish routine surveillance of surgical capacity, volume and results.

The checklist has been modified for its use in the UK taking into account the local healthcare system and practices and is now an essential tool in all surgical and radiological procedures.

Three separate checks are carried out:

1 *Sign in: before conduct of anaesthesia*

- Confirmation with patient of identity, site, procedure and consent.
- Confirmation of site marking.
- Anaesthetic machine and drugs checked.
- Allergies.
- Difficult airway and aspiration risk.
- Risk of >500 mL blood loss (7 mL/kg in paediatrics).

2 *Time out: prior to start of surgery*

- Introduction of team by name and role.
- Surgeon, anaesthetist and registered practitioner verbally confirm patient, site and procedure.
- Anticipated critical events.
- Surgeons review: any issues, blood loss, special requirements and investigations.
- Anaesthetic review: any concerns.
- Nurse/operating department practicioner review: any equipment and instrument issues.
- Surgical site infection bundle: antibiotic prophylaxis, warming, glycaemic control, hair removal.

3 *Sign out: before any member of the team leaves the operating room*

- Verbally confirm the name of the procedure recorded.
- Instruments, sharps and swab counts correct.

- Correct labelling of specimens.
- Equipment problems identified.
- Surgeon, anaesthetist and registered practitioner review key concerns for recovery and management.

The WHO also recommends briefing and debriefing at the beginning and end of the list to improve communication, situational awareness and team work.

Regional anaesthesia

This is anaesthesia that affects a part of the body. It is achieved by the use of local anaesthetics and other agents to remove the sensation of pain from a specific area. This can be as a single injection, by placement of a catheter for continuous infusion or by intermittent administration of the drug.

The indications include:

- sole anaesthetic technique for surgery in an awake or sedated patient
- in combination with general anaesthesia for intra- and postoperative pain relief
- for pain relief only, e.g. epidural analgesia during labour, chronic pain injections.

It can be broadly divided into:

- central/neuroaxial techniques: spinal, epidural, combined spinal epidural and caudal
- peripheral techniques, which are further subdivided into:
  - plexus blocks: brachial plexus, lumbar plexus blocks
  - individual nerve blocks: femoral, sciatic, median, ulnar nerves
- intravenous regional technique: Biers block (now rarely used)
- local techniques: wound infiltration, topical to mucous membranes, teeth.

Central/neuroaxial blocks

Spinal anaesthesia is achieved by injecting local anaesthetic in the intrathecal space (Figure 4.2). It is normally done at the L3–4 level (spinal cord terminates at L1–2 in adults) as a single injection but catheter techniques are also available. All lower abdominal, pelvic and lower limb surgery can be performed under a spinal block.

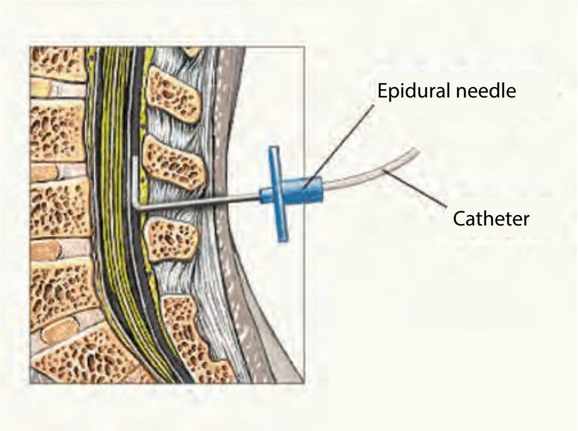


Figure 4.2 Sagittal view of the vertebrae and spinal cord, Tuohy needle in place ending in the epidural space and a thin spinal needle ending in the cerebrospinal fluid

Epidural anaesthesia is achieved by placement of a catheter in the extradural space. Normally its placement corresponds to the midpoint of the surgical scar, but it can be anywhere along the thoracic and lumbar spine. The block is normally not as dense as a spinal, but can cover higher dermatomes and so makes it ideal for pain relief in upper abdominal surgery.

Caudal anaesthesia involves insertion of a needle or cannula into the caudal space via the sacrococcygeal membrane and injection of approximately 20 mL of local anaesthetic. A catheter may also be passed. It is indicated for anaesthesia and pain relief for perineal and inguinal surgery.

The common complications of neuroaxial blocks include:

- failure: either immediate (e.g. catheter not in epidural space) or delayed (catheter migration)
- hypotension: intraoperatively or postoperatively owing to sympathetic block (common cause of hypotension in recovery).

Rare complications include:

- damage to structures: spinal cord, nerves, ligaments, blood vessels
- local anaesthetic toxicity owing to intravascular injection or peripheral absorption: agitation, fitting, loss of consciousness, severe arrhythmias, cardiac arrest
- high block: hypotension, bradycardia, difficulty in breathing/apnoea, reduced/loss of consciousness
- postdural puncture headache: due to cerebrospinal fluid leak through a tear in the dura mater (occurs in approximately 1:200 epidurals, rarer for spinals)
- epidural haematoma or abscess: needs urgent imaging and neurosurgical involvement.

The drugs used are:

- local anaesthetics: commonly bupivacaine, lidocaine, ropivocaine
- opioids: fentanyl, diamorphine, morphine
- other adjuncts: ketamine, clonidine, epinephrine (adrenaline), bicarbonate.

Peripheral nerve block techniques

Peripheral nerves and plexuses may be blocked by local anaesthetic injection guided by landmarks, nerve stimulator or ultrasound imaging. In general surgery, the most commonly used blocks are ilioinguinal nerve blocks, transversus abdominis plane blocks, paravertebral blocks and intercostal nerve blocks.

Transversus abdominis plane block

Local anaesthetic is injected in the space between the fascia of internal oblique and transverse abdominis muscles to block the anterior rami of the lower six (T7–L1) thoracic nerves. This provides good postoperative pain relief for a variety of procedures involving the lower abdomen down to the pubic region. It is commonly performed with ultrasound guidance for needle placement, and a catheter may be sited. This is particularly useful for unilateral abdominal surgery such as reversal of Hartmann's procedures, but may be performed bilaterally to cover lower abdominal laparotomy scars.

CASE STUDY 4

A 67-year-old man undergoes a total colectomy with an incision from the xiphisternum to pubic symphysis (dermatomes T6–12). An epidural catheter is inserted 5 cm into the epidural space and an infusion of 'low-dose mix' (bupivacaine 0.1% with 2 µg/mL fentanyl) is commenced at 10 mL/h. Block to cold sensation of T4–L1 is achieved and the patient is pain free, but still able to move and has no numbness. He is able to cough effectively and co-operate with physiotherapy.

Epidural catheters in general abdominal surgery

In recent years it has become commonplace to site an epidural to provide analgesia for abdominal procedures. When successful, this technique results in excellent pain relief with most patients reporting no pain at all. However, the postoperative management of epidural catheters can be difficult and should be done by the pain team and/or anaesthetist in charge of the patient.

Infusions used

Most commonly a combination of a dilute concentration of local anaesthetic (such as bupivacaine) is combined with an opiate (such as fentanyl) to provide a preferential block of unmyelinated C nerve fibres while sparing the other myelinated fibres. Consequently, a preferential block of pain is achieved while preserving touch and motor function. The infusion may be continuous, or a continuous background infusion with patient-controlled boluses (patient-controlled epidural analgesia).

Common problems

These include:

- Epidural never effective: usually due to the epidural catheter not being in the right place.
- Pain in the top of the wound: the epidural catheter may be placed too low (e.g. a lumbar epidural for an abdominal procedure) or the infusion rate may be inadequate.
- Pain in the bottom of the wound: the epidural catheter may be placed too high or the infusion rate may be inadequate.
- Epidural requiring repeated 'boluses' to achieve pain relief, with little or no block to cold sensation: the epidural catheter may have migrated, usually so that it is only partially in the epidural space, or not at all. Epidural boluses may result in temporary relief of pain even when the catheter is subcutaneous as systemic absorption of fentanyl is significant.
- Interruption in epidural infusion leads to breakthrough pain after 30 minutes to 1 hour.

Rare serious problems

These require prompt action:

- severe headache, worse on standing, accompanied by auras and nausea may indicate postdural puncture headache
- a dense block with paralysis during an infusion of dilute local anaesthetic may indicate an epidural haematoma or abscess and needs immediate imaging and intervention
- a rise in inflammatory markers in the absence of other signs of infection may indicate an epidural abscess; this is very rare.

Analgesia in the perioperative period

Principles of relief of pain

The WHO defines pain as 'an unpleasant sensory or emotional experience associated with actual or potential tissue damage, or described in terms of such damage'. It is important to note that pain is a subjective concept and patients' experience of pain after a similar procedure varies considerably.

World Health Organization pain ladder

The WHO pain ladder (http://www.who.int/cancer/palliative/painladder/en/) was originally described for cancer pain, but is now used as a basis for the treatment of many other types of pain, including postoperative pain. It consists of three steps:

1. mild pain: treated with non-opioids and adjuvant agents
2. moderate pain: treated with mild opioids in addition to the above
3. severe pain: treated with strong opioids, non-opioids and adjuvants.

Types of analgesic

These are categorized into (1) non-opioids, (2) mild opioids and (3) strong opioids:

- Non-opioid:
  - paracetamol (acetaminophen)
  - NSAIDs.
- Mild opioids:
  - codeine
  - tramadol.
- Strong opioids:
  - morphine
  - diamorphine
  - fentanyl
  - oxycodone.

Adjuvant agents may also be given, including:

- local anaesthetic infiltration
- regional analgesia
- gabapentin/pregabalin
- ketamine
- anxiolytics
- anti-inflammatory drugs (e.g. steroids).

The treatment of surgical pain should consist of a combination of agents appropriate for the expected level of pain. Regular administration of analgesics should be the mainstay of pain management with 'as required' agents available if necessary. The aim should be for the patient not to need to ask for pain relief, but for the drugs to be given regularly enough to relieve the pain.

Paracetamol (acetaminophen)

Paracetamol is a commonly used analgesic useful for mild to moderate pain. It can be given orally, rectally and by intravenous infusion. Its mechanism of action is still unclear despite much study, but it seems to be a highly selective cyclo-oxygenase (COX) 2 inhibitor, and may inhibit COX-3 in the brain. In addition, it may affect the brain cannabinoid system, producing pain relief. It has few side effects and few contraindications. Although it causes liver failure in overdose, it has been used in patients with liver impairment without adverse effects.

Non-steroidal anti-inflammatory drugs

These work by COX-1 and COX-2 inhibition, and have both analgesic and anti-inflammatory properties (unlike paracetamol, which has only analgesic properties). Because of their COX-1 inhibition they have a more significant side effect profile and contraindications. They work well, particularly for incision pain and bone pain.

The side effects of NSAIDs include:

- gastrointestinal ulceration
- exacerbation of asthma (most asthmatics are not affected by NSAID use)
- renal impairment
- platelet inhibition (especially aspirin).

Opioid analgesia

Opioids are strong analgesic drugs that act on opioid receptors – most on μ receptors in the brain and the spinal cord. They cause profound analgesia and are the mainstay of postoperative pain relief for moderate to severe pain.

Commonly used opioid analgesics include morphine, diamorphine, fentanyl and pethidine. Fentanyl may be given via any route, and morphine can be given orally or parenterally. Oral absorption of morphine is affected by first-pass metabolism and the bioavailability is under 50%.

The common side effects of opioid analgesia include:

- sedation
- respiratory depression
- nausea
- vomiting
- constipation
- urinary retention.

Patient-controlled analgesia

The principle behind patient-controlled analgesia is the ready supply of pain relief controlled by the patient themselves with some important safety features to prevent abuse and overdose. The details are as follows:

- the patient is attached to a pump containing the analgesic (usually morphine) via a dedicated intravenous cannula
- the pump delivers a dose of the drug (e.g. 1 mg morphine) every time the button is pressed
- there is a lockout period following each activation (e.g. 5 minutes) to prevent repeated dose delivery causing overdose
- the pump should only be activated by the patient themselves (or trained nurse in children, for example) – never by relatives
- a background infusion may be used to reduce the number of times a patient has to activate the pump.

Patient-controlled analgesia works best when prescribed with coanalgesics such as paracetamol and NSAIDs.

GUIDE TO FURTHER READING

Allman K, Wilson I. *Oxford Handbook of Anaesthesia*, 2nd edn. Oxford, UK: Oxford University Press, 2011.

National Institute for Health and Clinical Excellence. *Chronic Obstructive Pulmonary Disease*. Clinical Guideline 101. London, UK: NICE, 2010. Available from: http://guidance.nice.org.uk/CG101.

Smith T, Pinnock C, Lin T, Jones R. *Fundamentals of Anaesthesia*, 3rd edn. Cambridge, UK: Cambridge University Press, 2009.

Yentis SM, Ip J, Hirsch NP, Smith GB. *Anaesthesia and Intensive Care A-Z: An Encyclopedia of Principles and Practice*. Edinburgh, UK: Churchill Livingstone, 2009.

Weblinks

www.aagbi.org/publications/publications-guidelines

www.frca.co.uk/

www.who.int/cancer/palliative/painladder/en/

www.who.int/patientsafety/safesurgery/en/

CHAPTER 5

Patient-reported outcomes to evaluate surgery

JANE M. BLAZEBY

Introduction

Evaluation of surgery can be defined as a rigorous assessment of the advantages and disadvantages of interventions both in the short and long term and this process is important for surgeons, patients and health service providers. Evaluation may involve audit or research methods, and both can improve the quality of care for patients because capturing data about outcomes, and making information available to key stakeholders means that action and collective responsibility can be taken for the results. Whatever the type of study design used to evaluate surgery it will be necessary to choose, measure and report outcomes which are defined as endpoints that gauge the effectiveness of an intervention. Both the design of the study *and* the selection and measurement of the particular outcomes are, therefore, of critical importance to inform practice.

Traditionally outcomes used to evaluate surgical interventions have been those selected by surgeons regarding complications, and these have predominantly been morbidities occurring within the same hospital stay as the procedure itself. Outcomes of utmost concern for surgeons usually relate to perceived measures of technical performance, such as rates of anastomotic leakage following bowel resection, rates of arterial occlusion in peripheral vascular surgery or rates of bleeding after cardiac surgery. There has also been a focus on reporting longer term surgical outcomes, including how surgery may affect functions (e.g. bowel function following subtotal colectomy) or how successful surgery has been in removing the original problem, e.g. 5 year survival rates after surgery for cancer or recurrence of reflux problems following fundoplication. Although measuring and understanding short- and long-term complications will always be of paramount importance to surgeons, the assessment of outcomes of interest to other stakeholders will provide a more comprehensive evaluation of interventions to inform decision-making. Over the past three decades there has been a growing interest in measuring patient-reported outcomes (PROs) to provide data to use in shared decision-making consultations between surgeons and patients.

PROs assess any aspect of health that is reported by patients themselves and that is not interpreted by an observer, and patient-reported outcome measures (PROMs) are the instruments designed to assess a PRO. Understanding how PROs may be used to evaluate surgery is the focus of this chapter, which will define PROs, PROMs and the glossary of terms used in this field. The chapter will provide an explanation of the importance of PROs to surgical practice and a summary of the different types of PROMs, and it will outline how to design a robust study to measure PROs to evaluate surgery. It will illustrate these issues with surgical case studies. Areas for future research in this field will also be considered.

What are patient-reported outcomes?

A PRO is an outcome that assesses any aspect of a patient's health that comes directly from the patient, without the interpretation of the patient's responses by a physician or anyone else. Historically, few scientific evaluations of surgery made an attempt to measure patients' views, and only doctors' or nurses' assessment of patients' symptoms, functional health status or overall quality of life (QOL) were sometimes reported. One of the first instruments that did this, thereby at least broadening outcome measurement beyond clinical measurement of complications, was the Karnofsky Performance Scale proposed for use in 1947. This scale, and several other measures of performance that were (and are still) widely used, asked the clinician to rate from 0 to 100 how the patient's general performance appeared to them. Zero indicated dead and 100 was normal health. The World Health Organization's measure of performance is nowadays used in clinical practice, particularly in oncology. In surgical studies, formal measures of performance have been less widely used, although it is standard for surgeons to make judgements about general health when deciding if a patient is sufficiently fit for a procedure. In the one multicentre audit examining outcomes of gastric cancer surgery, surgeons' judgements of patients' general fitness were used to select patients, and analyses showed that these were prognostic. Using an observer assessment of general health, however, without a definition means that it is not easily reproducible and may be criticized as being invalid. The problem with measures of patients' health made by anyone other than the patient themselves is that they may be incorrect or biased, being an over- or underestimate of the patient's true functional ability. Another study compared several aspects of QOL, including dysphagia scores reported by patients, by carers or by a clinician. Both over- and underestimates were observed and there was only poor to moderate agreement about most aspects of health. Observers may misjudge symptoms, functional aspects of health and make inaccurate global assessments. It is therefore necessary that every effort is made to obtain the information from patients themselves so that an evaluation of the impact of surgery of PROs can be made.

A PRO may refer to a single dimension of health, a global health assessment or it may include a multidimensional assessment of several different aspects of health. A single dimension of health means a report of a single symptom or function, such as a self-reported pain score or a self-reported measure of mobility. A global health assessment means that the patient responds to a question that asks them to rate 'overall health' or 'overall quality of life', and a multidimensional PRO measures more than one aspect of health (e.g. pain *and* mobility). Box 5.1 provides a glossary of definitions of terms commonly used in PRO studies. Other commonly used words include 'health status', 'quality of life' and 'health-related quality of life' (HRQL/HRQoL). Each of these is a PRO only if the assessment is performed by the patients themselves. For example, a doctor may describe and report a patient's QOL, but this is an observer assessment of QOL and not a PRO. However, it is possible for patients to self-report their overall QOL and this would be classified as a PRO.

BOX 5.1 Definitions of terms commonly used in patient-reported outcome studies

- **Patient-reported outcome (PRO)** is an outcome that assesses any aspect of a patient's health that comes directly from the patient, without the interpretation of the patient's responses by a physician or anyone else
- **Patient-reported outcome measure (PROM)** is an instrument that is used to assess any PRO
- **Quality of life (QOL)** is an ill-defined term that may have components of happiness or satisfaction
- **Health-related quality of life (HRQL, HRQoL)** is also ill defined, but it is generally agreed that it can include several aspects of health, such as physical, emotional and social functioning and symptoms
- **An item** is a single question that may assess any aspect (domain) of QOL/HRQL or any PRO
- **A scale** is two or more items in a questionnaire that assess the same health domain
- **A domain** is an aspect of health (e.g. pain, physical wellbeing); it may be assessed by a single item or a scale
- **A global assessment of QOL or HRQL/HRQoL** is an item or scale that assesses generic health or global QOL issues (e.g. how is your overall quality of life?)
- **Patient-reported experience** is a measure of the process of care received by the patients, such as levels of satisfaction with information provision, satisfaction with access to car-parking facilities and experiences waiting in the hospital
- **Patient-reported experience measure (PREM)** is an instrument that is used to assess any PRE

There are no strict internationally agreed definitions of QOL and HRQL, and in layman's terms both mean abstract notions of happiness or satisfaction that may be related to current or past events and individual values. Within a health services research context, however, both QOL and HRQL are terms that are interchangeable and refer to the patient's perception of health, including single and multidimensional components. Because of the potential for confusion with this terminology, it is recommended that, in any evaluation of surgery with PROs, a definition of the health domains of interest is provided in the protocol and report so that readers are aware exactly which aspects of health are being assessed. It is also recommended that clear descriptions of who made the assessment are provided and whether it is truly a PRO or an observer assessment of the patient's health.

It is important to distinguish between a PRO and a patient-reported experience (PRE). A PRE has been referred to as a measure of the process of care received by the patients, such as levels of satisfaction with information provision, satisfaction with access to car-parking facilities and experiences waiting in the hospital. The role of assessing PREs is different from measuring PROs and they will be considered separately.

Patient-reported outcome measures

PROMs or patient-reported experience measures (PREMs) are the instruments used to assess PROs and PREs, respectively. These are most frequently paper-based questionnaires, although it is likely that future methods for capturing PRO data will include electronic means. Whether paper-based questionnaires or electronic methods are used to measure PROs it is critical to use tools that are tested for reliability and validity. There are several types of PROMs. Some of the instruments that are more widely used are described with an outline of their scoring procedures and any constructed scales.

Generic instruments

Some questionnaires are designed for use in any patient population to capture the generic aspects of health. They commonly include scales assessing physical, social, role and emotional function. Some of the more widely used tools have also been updated to include short-form versions to allow a swift assessment of the important generic aspects of health. Table 5.1 summarizes key aspects of commonly used questionnaires including the SF-36 and the EQ-5D. These are available in a range of languages and have been widely used in

Table 5.1 Commonly used generic and disease-specific patient-reported outcome measures

Name of instrument	Number of items and scales	Health domains	Time frame	Scoring system
Medical Outcomes Study Short Form 36 (SF-36)	36 questions 8 health domains	Physical function Role physical Bodily pain General health Vitality Social function Role emotional Mental health	The past 4 weeks Some refer to 'now'	8 scales can be scored or two summary measures, physical and mental health, can be created
EuroQoL (EQ-5D)	5 questions	Mobility Self-care Usual activities Pain/discomfort Anxiety/depression	Today	A single overall index is created
European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)	30 questions 9 scales 6 single items	Physical function Role function Cognitive function Social function Emotional function Global health scale Financial impact (9 symptoms, including pain, fatigue, nausea and vomiting, dyspnoea, anorexia, insomnia, constipation, diarrhoea)	The past week	Each scale and item is scored individually
Functional Assessment of Chronic Illness Therapy (FACIT)	27 questions 4 health domains	Physical wellbeing Social wellbeing Emotional wellbeing Functional wellbeing	The past 7 days	4 scales and an overall summary score
Multidimensional Fatigue Index (MFI)	20 questions 5 domains	General fatigue Physical fatigue Mental fatigue Reduced motivation Reduced activity	'Been feeling lately'	5 scales and an overall summary score
Barthel Index of disability (BI)	10 questions 1 domain	Functional dependency	Not specified	One overall score
Katz Activity of Daily Living scale (Katz ADL)	6 questions 1 domain	Independence	Not specified	One overall score

clinical trials. Generic instruments have the benefit of covering core aspects of health and they allow for cross-study and cross-population comparisons to be made. Disadvantages are that they may not be sufficiently sensitive to capture symptoms or problems peculiar to a given operation or disease. The SEIQoL and the Patient Generated Index are examples of generic health measures that allow the participants to nominate areas of health that are important individually to them, and the instruments examine the current level of functioning within the selected areas. Although these seem appealing, they are less easy to use and score, and making comparisons between study groups is less intuitive because of the knowledge that different components of health make up for the summary scores that are produced.

Disease-specific patient-reported outcome measures

Disease-specific measures have been developed to detect the issues of relevance to patients with particular conditions. Examples of measures developed for cancer are given below, including the EORTC QLQ-C30 and the FACIT questionnaires. Each one assessed issues that are relevant to a wide range of patients with cancer. Both may be supplemented by disease- or treatment-specific modules that will address aspects of treatment or the disease relevant to a patient population. Both questionnaires relate to the past week, and they are extensively tested for reliability and validity in many different populations of patients with cancer. Both have many different modules available to use for specific diagnoses or treatments and studies have demonstrated the value of these tools after cancer surgery.

Instruments for specific patient-reported outcomes

The above instruments all measure PROs of several aspects of health/QOL and some include questions that ask patients to make global assessments. There are situations when a very specific aspect of health is the main outcome and then instruments may be designed to explore issues in much greater depth. Since such instruments measure just one aspect of health/QOL, they are often combined with a generic measure to ensure that the assessment is multidimensional. Two example questionnaires are described in Table 5.1, the Multidimensional Fatigue Index (MFI) and indices to measure activities of daily living (ADLs), the Barthel Index of Disability (BI) and the Katz ADL. The MFI measures fatigue including physical and mental aspects and the BI is an ADL index that measures functional dependency before and after treatment. The Katz ADL measures independence. Measurement of self-reported fatigue and activity may be particularly valuable to use in surgical trials when a PRO assessment of postoperative recovery is required.

Features to consider when selecting a patient-reported outcome measure

In addition to selecting the type of PROM (generic, disease or domain specific) it is necessary to examine the practical aspects of questionnaires and details of their measurement properties. Tips are summarized in Table 5.2. Initial selection will depend upon the type and content of the questionnaire and examination of the scales and questions in the instrument in detail. One method that is useful to look at content, practicality and appropriateness of the questionnaires is to ask a patient to complete the questionnaire while you observe them and their understanding of each item. During this exercise the time taken to complete the questionnaire can be checked, as ensuring that the questionnaires are completed within 20 minutes will optimize response rates and data completeness. Many questionnaires have been formally translated, and if the patient population is multicultural this will be essential to get results that are generalizable. Developers of questionnaires may require a formal licence and a user agreement, sometimes with a fee for the questionnaire to be used. The details are available on the questionnaire websites usually, and may be study or user specific.

A detailed examination of the measurement (psychometric properties) of the questionnaire is important unless a very widely validated and used measure is selected for which you can be confident that these are correct. Ensuring that internal consistency, construct and criterion validity, and test–retest properties are published and of acceptable standards means that the PROM will assess what it is reported to and that it will produce repeatable results. In addition, information about the sensitivity of the questionnaire to changes in health status over time is relevant to longitudinal studies or clinical trials. Traditional psychometric methods include summated scales and factor analysis techniques and it is also possible to examine measurement properties using item response theory models. Both require appropriate expertise.

Table 5.2 Points to consider when selecting a patient-reported outcome measure (PROM)

Questions to ask	How to find out the answer
Is it a PROM?	Check that it is completed by the patient
The content of the questionnaire	
What does the questionnaire measure?	Read the questions
What scales and items are included?	Look at the scoring for the scales/items
Is it relevant to the targeted patients?	Observe a patient completing the questionnaire
Practical issues	
How long does it take to complete?	Complete it and time yourself, examine evidence
Is it available in other languages?	Look at published evidence
Is there a user's agreement?	Contact the developers/check publications
Is there a cost for using the PROM?	Check the developer's details
How is the questionnaire scored?	Look at publications and get the scoring manual
Where are copies of the PROM obtained?	Check questionnaire instructions/read questions
Content validity	
What informed the content of the PROM?	Check the developmental publications
Were patient interviews performed?	Were qualitative patient interviews performed?
Was the literature searched?	Was a systematic review of PROs performed?
Were health professionals views included?	Were professional interviews undertaken?
Reliability	
How were the PROM scales formed?	Examine the factor analyses or multitrait scaling
Are the Cronbach α scores satisfactory?	Scores above 0.7 are good
Has it be tested for reproducibility over time?	A stable instrument has high correlations
Construct and criterion validity	
What domains/clinical outcomes has it been compared with?	Check for comparisons with validated PROMs
Are there correlations with health domains?	Examine publications
How were the scales developed?	Examine validity data in publications
	Examine factor analyses and multitrait scaling
Sensitivity to changes in health status	
Does it cover the relevant time?	Read the questionnaire introduction and the response categories carefully
Does it capture extreme responses? (e.g. severe pain as well as absence of pain)	

PRO, patient-reported outcome.

Core outcome sets for the evaluation of surgery

One of the problems with PRO assessment is the myriad available questionnaires and scales (PROMs) to use. This means that studies often report up to 25 different PROs and that data are difficult to synthesize (because different measures are used). It means that studies may suffer from outcome reporting bias, in which data are presented and published because of statistically significant findings (omitting non-significant results) rather than publishing the results of prespecified hypotheses. These methodological problems are also common in clinical studies, because of the multiple, ill-defined methods used to assess recovery and outcomes of surgery. A potential solution to the problem of outcome reporting is to define core outcome sets for each surgical condition or procedure. A core outcome set is an agreed set of outcomes (usually about 7–10) that are reported in all studies, as a minimum, evaluating a particular condition or intervention. The systematic use of core outcome sets will improve direct and indirect comparisons of trials, data synthesis in meta-analyses, audit of hospitals or surgeons' performance and also reduce outcome reporting bias.

The OMERACT (Outcomes Measures in Rheumatoid Arthritis Clinical Trials) initiative has pioneered this approach, initially developing a core outcome set for rheumatoid arthritis but more recently expanding this to condition- and disease-specific measures. Core outcome sets are developed using an iterative data-driven scientific approach that includes surveying views of key stakeholders to reduce an initial long list of all potential outcomes to a core list that is agreed by reaching consensus. OMERACT initially developed a core clinical outcome set, with seven items selected that were predominantly chosen by rheumatologists. This was subsequently expanded after work with patient groups further identified core 'PRO' sets, and one of the PRO outcomes, fatigue, was subsequently included in the core clinical outcome set. This approach is very likely to improve outcome measurement and reporting in trials and clinical audit and the development of core outcome sets for surgical procedures, including PROs, is strongly recommended. The COMET (Core Outcomes in Effectiveness Trials) (http://www.comet-initiative.org/) initiative brings together researchers interested in the development and application of agreed standardized sets of outcomes. The current sets of outcomes available are collated on the website, and the initiative aims to collate and stimulate relevant resources, both applied and methodological, as well as facilitating exchange of ideas and information, and fostering methodological research in this area.

Summary

Currently, there are a wide range of PROMs available, although as the science in this area develops and the clinical application of PROs is better understood it is likely that these will be reduced. Key aspects of PROMs are to ensure that they are designed to be completed by patients themselves, that they measure the dimensions of interest to the study and that they cover the time frame of interest with response categories that are appropriate. Methods to score each instrument and present clinically meaningful results are developed and available for validated tools. The interpretation of the PRO data alongside clinical outcomes is especially important for results to be used in practice and be meaningful to patients and clinicians. The potential development and identification of core outcomes sets for particular diseases and interventions that include PROs is a practical method for ensuring that outcomes can be synthesized and compared between studies, and it may reduce the problem of outcome reporting bias.

Why measuring patient-reported outcomes is important in the evaluation of surgery

Using patient-reported outcomes in clinical practice

The main roles for measuring PROs in addition to clinical outcomes to evaluate surgery is to use the data from well-designed studies to (1) meet patients' information needs, (2) inform disclosure for consent for surgery and (3) inform clinical and health policy decision-making. These overlapping areas, which all involve the surgeon knowing and understanding PROs from well-designed studies evaluating surgical procedures, are considered in detail below. The PRO data may come from studies examining a single surgical intervention (by describing how the operation impacts on short-term and longer term PROs), or data may come from randomized or non-randomized comparative studies.

Patient-reported outcome data to meet patients' information needs

The information needs of patients scheduled for surgery differ among patients and change over time. Evidence shows, however, that most patients report that they want to know as much information as possible, although in some situations they may be overwhelmed by the sheer amount of information provided by clinicians, the mass media and the internet. In addition to patients wanting information, the family may also be requesting data, and dealing with these issues can be difficult. The type of information that patients want to know includes clinical details about risks and chances of surgery being successful, and patients and families want information about psychosocial and emotional issues and how the patient will feel after the operation. Accurate provision of such information by surgeons during consultations with patients requires that well-designed studies with clinical outcomes and PROs have been performed, that surgeons are aware of these outcome data and that they are able to communicate it effectively.

Communication of risks and benefits of surgical procedures is recommended before consent can be obtained for surgery, and, in the UK, guidance from the General Medical Council also recommends that information provision includes data about how surgery will affect lifestyle and factors that are important to the patients themselves. It is these latter details that require studies to be performed with reliable and valid PROMs, otherwise surgeons will not be aware of how surgery affects psychosocial wellbeing for example, or patients' ability to function.

Some types of surgery have used PROMs in well-designed studies and produced useful data that can be used in consultations and in written information sheets to provide patients and families with relevant details to meet information needs. At present, methods to communicate PRO data are not standardized in practice, although patients may have high levels of understanding of bar charts or pictographs to communicate multidimensional PROs. Explaining risks and rates of clinical outcomes alongside PRO data may be challenging, however, especially if surgery is major. Decision aids have been developed to help clinicians to communicate complex information to patients and data within these tools is usually derived from well-designed studies. Decision aids may be in the form of booklets, prompt sheets, video recordings and interactive web-based formats, which can help patients understand data and lead to a more active role in decision-making. There is evidence that decision aids increase patients' knowledge about options, lower decisional conflict and reduce the proportion of patients remaining undecided about treatment.

Standards for information disclosure in informed consent for surgery

It has already been stated that information needs of patients vary and change over time and, although many patients want as much information as possible, some patients prefer to leave the decisions to the surgeons – taking a passive role in the consultation. Surgeons, therefore, need to make efforts to elicit the information that patients may wish to have and to provide data about clinical outcomes and PROs and information about the process of care to meet each patient's needs. In situations in which patients do not want to have information, there is also the requirement for surgeons to provide at least sufficient information for patients to understand what they are agreeing to undergo and provide 'effective' consent. If patients are unaware of any of the consequences of surgery, they may not realize what they really did want to know in more detail. This highlights the differences between societies' rules governing 'effective consent' and genuine autonomous consent from a patient who truly understands what they are signing up for. Legal consent, for example, may focus on doctors' responsibilities to disclose information, rather than ensuring that patients understand and that they have received information of importance to their values.

There are three theories of informed consent that are relevant to surgeons informing patients of the outcomes of surgery: the professional practice standard (also called the prudent or reasonable doctor standard), the reasonable person standard (or prudent/reasonable patient standard) and the subjective standard. From a professional viewpoint it is theorized that only doctors have the necessary expertise and experience to decide what information patients need to undergo surgery. It is considered that doctors are acting in the best interests of patients (beneficence). Provided that a responsible body of medical opinion agrees on what information needs to be communicated as part of consent, then that is the minimum information required to avoid being negligent. This theory may be criticized, however, and it only views information disclosure from a medical viewpoint, meaning that outcomes of importance to the patients may not be disclosed, overriding principles of autonomy. The reasonable person standard has therefore gained acceptance in over half the US states and in the UK. In this standard, information is deemed important according to what a hypothetical 'reasonable person' might expect. Typically, this information includes the nature and purpose of the treatment, risks, potential benefits and available alternatives. The reasonable person standard confers some considerable ethical advantages: it shifts the determination of what information is important away from the doctor and moves towards ideas of 'autonomous authorization'. It does, however, risk being too patient focused and not allowing patients to gain sufficient understanding of the medical outcomes of an intervention, although it may be argued that there is no moral distinction between reasons why information is important. According to this theory, information that influences individuals' decision-making may be equally important as that which does not. There is therefore another standard that needs consideration when considering information disclosure. In this standard, the information that needs to be disclosed for legal consent is taken on a case-by-case basis. This is because of the subjective personal values that patients use in defining what is 'important'. This standard is not used in the UK. Problems centre on the legal credibility of patients' retrospective opinions, particularly when patients perceive they have been wronged. In addition, it has been considered unreasonable for a doctor to carry out an exhaustive character analysis to determine each patient's relevant needs. The subjective standard may be seen as the preferable moral standard, but, because of the subjectivity, it has been described as insufficient for law.

Patient-reported outcome data to inform decision-making

Between patients and surgeons

Information exchanged during consultations between surgeons and patients in which decisions to undergo surgery are made is necessary to allow patients to reach an informed decision. Surgeons are ethically and legally required to provide patients with *sufficient* information to ensure that competent and autonomous patients have *substantial* understanding. Information regarding the disease, its natural history, how it may be treated (including surgical and non-surgical alternative treatments) and short- and long-term clinical outcomes and PROs may be communicated. The framework in which information is exchanged is recommended to be one of mutuality, 'shared decision-making', in which both the patient and the surgeon engage in the conversation and there is joint negotiation of values and information that is considered to be important to both parties. Typically, surgeons know the clinical consequences of surgery and rates of these happening, and they can make sensible estimates of these occurring individually based upon risk assessments of the patient's general health, comorbidities and nature of the disease severity. Surgeons, however, are not often aware of how most procedures may affect PROs and they may not be aware of the patient's individual values about different sorts of outcomes. For decision-making, therefore, to be 'shared', it is necessary for a surgeon to be aware of the PROs of each procedure and to elicit the patient's values during the consultation. These principles are supported by most surgical organizations, and in the Royal College of Surgeons of England publication *Good Surgical Practice* (which is endorsed by 12 other organizations) it is stated that:

> A surgeon must listen to and respect the views of patients and their supporters; recognise and respect the varying needs of patients for information and explanation; insist that time is available for a detailed explanation of the clinical problem and the treatment options and encourage patients to discuss the problem.

Therefore, during consultations in which decisions are made to undergo surgery, it is necessary to provide information of relevance to the patient as well as clinical data. Figure 5.1 illustrates how this may work in practice.

Between health policy-makers

PRO data can also be used to inform health policy-makers during decisions in which resource allocation is agreed. This process varies between countries, and the UK model will be described here. In the UK, the National Institute for Health and Clinical Excellence (NICE) provides guidance, sets quality standards and manages a national database to improve people's health and prevent and treat ill health. It makes recommendations to the National Health Service,

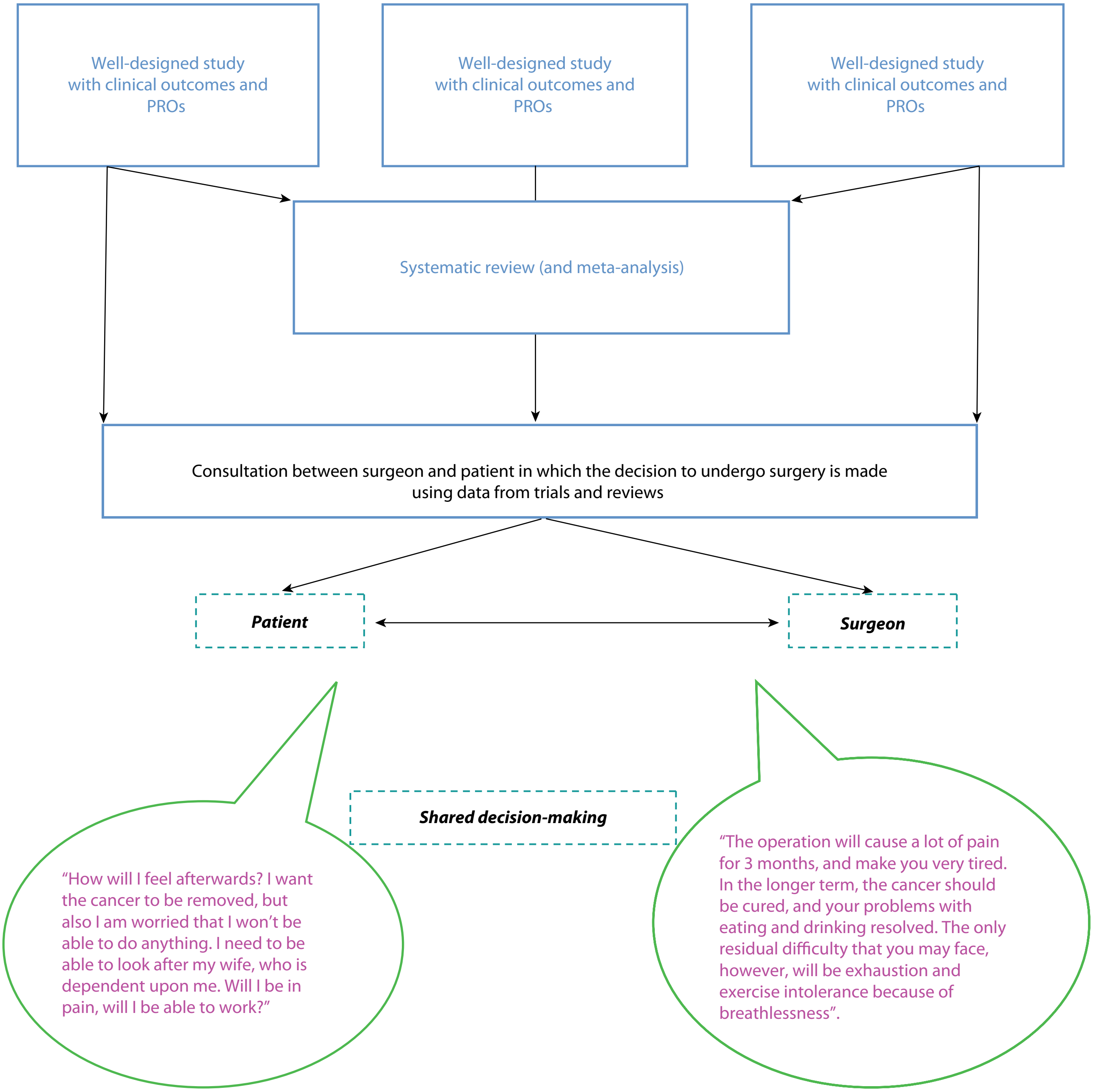

Well-designed study with clinical outcomes and PROs
Well-designed study with clinical outcomes and PROs
Well-designed study with clinical outcomes and PROs
Systematic review (and meta-analysis)
Consultation between surgeon and patient in which the decision to undergo surgery is made using data from trials and reviews
Patient
Surgeon
Shared decision-making
"How will I feel afterwards? I want the cancer to be removed, but also I am worried that I won't be able to do anything. I need to be able to look after my wife, who is dependent upon me. Will I be in pain, will I be able to work?"
"The operation will cause a lot of pain for 3 months, and make you very tired. In the longer term, the cancer should be cured, and your problems with eating and drinking resolved. The only residual difficulty that you may face, however, will be exhaustion and exercise intolerance because of breathlessness".

Figure 5.1 How the data from studies with patient-reported outcomes may be used to inform practice.

to local authorities and to other organizations about new and existing surgical procedures and it aims to improve peoples' health as well as keeping costs to a minimum. One of the key indicators used by NICE is a PROM combined with a quantitative outcome, the quality-adjusted life year (QALY). This allows decisions on cost-effectiveness, usually measured in terms of incremental cost per QALY gained, to be made. A QALY places a weight on time lived by patients in different health states. A year of perfect health is worth 1, and a year of less than perfect health is worth less than 1. Death is considered to be equivalent to zero; however, some health states may be considered worse than death and have negative scores. Using QALYs provides a common currency to assess the extent of the benefits gained from a variety of interventions. When combined with the costs of providing the interventions, cost–utility ratios result; these indicate the additional costs required to generate a year of perfect health (1 QALY). Comparisons can also be made between interventions, and priorities can be established based on those interventions that are relatively inexpensive (low cost per QALY) and those that are relatively expensive (high cost per QALY).

The most commonly used PROM for the measurement and QOL when calculating QALYs is the EQ-5D. The EQ-5D is widely used and has been validated in many different patient populations. It has been designed so that people can describe the extent to which they have a problem in each of the five dimensions of health: mobility, ability to self-care, ability to undertake usual activities, pain and discomfort, and anxiety and depression. The responses to the EQ-5D allow 245 health states to be represented. The valuation of the different health states was calculated by a random sample of 3000 people in the UK examining these states and this allows the PRO data to be mapped to utility scores and used to form a QALY. Although the EQ-5D is the most widely used instrument for this purpose, some of the other validated measures of QOL have more recently been mapped into QALYs.

In surgical settings, gaining high-quality data about QALYs is unusual, however, as few trials and prospective studies are performed using this measure, and even fewer have made efforts to systematically capture and report PROs. The next section describes how to design studies with PROs that produce valid data to inform practice.

Using patient-reported outcomes in research

Within a research context there are other roles of PROs that may be relevant to clinical practice in the future. These include using PROs in outpatient consultations to collect detailed information about the range of problems that a patient may face and to use this in real time to facilitate communication in the clinic. PROs may be used as prognostic markers to predict survival after surgery for cancer and it is possible that PROs may be used to monitor patients after surgery for cancer and replace standard clinical follow-up, provided there are good systems in place to respond to PRO data. Although these areas are of interest, a lot more research and development is still required to be certain of the validity of PROs in these situations.

Using patient-reported outcome measures in clinical practice

There is a growing interest in routinely collecting PROs as part of care for individual cancer patients. The purpose of this is to capture patients' perspectives on the impact of their disease and treatment on their own health, including their symptoms, self-reported functions and their assessment of overall health. Feedback of PRO data in real time to clinicians responsible for their care has the potential to help detect unmet needs or problems, theoretically leading to better control and monitoring of the relevant issues. This may also prompt health professionals to discuss issues that may not have been raised by the patients in the clinic and facilitate better communication and care. There is some evidence to show the positive effects of this approach to processes of care (e.g. doctor–patient communication), but less evidence to show that this improves patients' outcomes *per se*. Reviews have summarized these issues and shown how this intervention is highly complex and dependent upon clinician and patient behaviour as well as the nature of the PROMs themselves. They also show how organizational changes are required to facilitate this process and electronic data collection and systems for feedback of the data to clinicians are needed. More recent work shows that doctors may receive PRO data about symptoms more readily than data about functions. Such work has predominantly been carried out in cancer patients and with oncologists. Further work evaluating the use of PRO data in real time in surgical settings is still needed.

The prognostic value of patient-reported outcomes

The predictive value of PROs has been examined in several situations and summarized in reviews. Many of the papers have not rigorously adjusted the multivariable models for known biological or clinical indicators, but those that have also show how some domains of HRQL are independently predictive of survival in many settings in surgical oncology. The underlying reasons for these observations remain unknown, whether patients with clinically unrecognized conditions feel more ill and report this in terms of PROs before the disease is diagnosed or its severity known is one possible explanation, or it may be that better PROs are associated with a healthier lifestyle and longevity. This is an area where further research is required.

Using patient-reported outcomes to monitor and follow-up patients

The use of PROs to monitor and follow-up patients is very plausible because it may reduce the need for outcome consultations, where a simple questionnaire can suffice to detect persistent or new issues. This has been examined in several non-surgical settings and results show that this is an expensive process with little impact on clinical practice. Whether following surgical interventions this may have better results and whether the use of electronic devices (e.g. by mobile phone) will make this cheaper and more readily usable in the health services is unknown and warrants further research.

Designing studies with patient-reported outcomes

The design of a PRO study that will yield data to inform practice is essential to ensure that outcomes are valid, that the results are believed and that they can be implemented. In addition to careful study design, the reporting of the study is critical and these two concepts are linked. A well-designed study should lead to a high-quality report, although even a weakly designed study can be well reported, allowing the reader to judge the validity of the study findings. A summary of the key things to consider during study design and reporting is given in Table 5.3. Several papers have recommended methods to ensure that PRO studies yield valid data in randomized clinical trials and the key issues relevant to trials and other studies are described below.

Selecting the time points to measure patient-reported outcomes before and after treatment

Selection of the relevant time to assess PROs will depend upon the study objectives and whether PROs are primary or secondary endpoints. This can be achieved when the study rationale is considered, and the PRO domain(s) or interest selected. It is recommended that the specific nature of the anticipated effects of the intervention and the expected time for them to be exerted be considered in detail, because this will influence which PRO is selected to be measured, which tool is used to measure it and the timing and frequency of its administration. For example, in a trial of minimal access gastrointestinal surgery versus an open procedure, a PRO measure assessing pain or physical function may be important to distinguish differences between the two interventions. This, however, will need to be assessed within a few days or weeks of surgery, or the differences between the two procedures may be less apparent if an assessment is performed 4 or 6 weeks after surgery.

In addition to giving consideration to the timing of the postoperative PRO assessment it is recommended that a pre-treatment assessment of PROs is performed to provide baseline data. The boundaries of the time windows around the baseline assessment need specification (e.g. within 4 weeks before surgery or within 4 weeks before randomization). Capturing PRO before treatment can be difficult but it is essential because it is extremely difficult for patients who have undergone surgery to look back to reflect about their symptoms and functional ability before the operation.

Ensuring that the right domain of health is measured in a study, with the best instrument, will require discussion within a multidisciplinary team and inclusion of a patient (user representative) in addition to surgeons and methodologists with experience and knowledge of PROMs.

Choice of patient-reported outcome measure

The choice of instrument will depend upon what PROs require measurement and features of the study population, including clinical and sociodemographic details, such as the type of intervention being evaluated, level of education, gender, culture, general health, diagnosis, native language and age range. Assessment of the measurement properties of the PROM is critical and the instrument should be considered for reliability, validity and responsiveness to change. The citation for the instrument development and psychometric testing should be checked with advice from an expert, and the actual questionnaire scrutinized by the study team (details above). Consideration can be given to how long it takes for the questionnaire to be completed, the time frame of each question (asking about problems in the past week or past month for example) and the domains of health

Table 5.3 Issues to consider when designing patient-reported outcome (PRO) studies

Study population	Consider clinical (e.g. prognosis) features Consider demographic features (e.g. language, literacy, age)
Length and content of questionnaire	Check content and length for study population
Optimizing response rates to questionnaire	Select a few assessment points Ensure that all staff are familiar with the protocol Use central questionnaire administration Chase missing responses regularly Provide feedback to units
Dealing with missing data	Establish clear practice to follow in the event of missing assessments Establish clear practice to follow in the event of missing items/pages Document reasons for missing data
Analysing missing data	Document the amount and reasons for missing data and consider causes for bias Impute missing items according to developer's instructions
PRO data analyses	Document missing data Follow scoring instructions of the questionnaire Initially examine descriptive data from scales Be aware of values that are considered to be minimally important clinical changes in PROs Be aware of multiple testing

covered. Table 5.1 shows some of the most widely used validated measures available to assess PROs in surgery.

Determining the health-related quality of life sample size

In randomized clinical trials in which the PRO is a primary endpoint, it will be necessary to consider the effect size of each intervention to determine the number of participants and the expected differences in effect size between each arm of the trial, as well as the required power for the trial. It is also necessary to consider the number of PRO assessments and the proposed analyses and the available information on the underlying assumptions of expected benefit. In general, PROs are secondary endpoints in trials and therefore they do not influence the sample size calculation. Where they are primary endpoints, consideration for non-response and loss to follow-up is also necessary. Although trials measuring PROs often include numerous patient reported outcomes measures (PROMs), it is recommended that a maximum of four or five outcomes *a priori* be included in formal analyses, taking into account the effect of multiple significance testing.

Missing data

Difficulties with obtaining good questionnaire compliance are commonplace in many studies and this can invalidate the results. The most frequent reason for missing questionnaire data is administrative failure and patients are simply not approached and asked to complete assessments. Missing data may also occur because patients are too ill (more common in cancer studies), and about 5% of data are missing because patients refuse to participate. Ensuring high questionnaire compliance is very important, and levels of over 80% should be aimed for to avoid response bias. A number of strategies can be enlisted to minimize missing PRO data. High-quality study design with clear and minimal PRO assessments that are straightforward to complete is one of the most important issues. Likewise, a well-designed study will be appropriately funded and ensuring that staff are familiar with the protocol and understand how to access patients is essential. Utilizing postal administration of questionnaires from a central trials team (after the patient has been recruited locally) is an efficient means of ensuring that questionnaires are administered on time and that missing responses are chased up appropriately.

For surgical trials, it is recommended that an in-built pilot study or a stand-alone feasibility trial be carried out to examine the best methods to collect PRO data and to allow improved organization and preparation for unforeseen difficulties in data collection. Patient-based sources of missing data may be tackled by choosing an appropriate instrument and by paying attention to participants' motivation by providing a clear explanation of the reasons of why, when and how PRO assessments will be made, whom they may contact for help and what will happen to the data they provide (an information sheet that details confidentiality and data dissemination is typically required by most ethics committees). Proxy completion by a carer, clinician or significant other may be considered in the event of missing data as a result of participants' inability to complete the questionnaire. This should be considered prior to the start of the trial if it is anticipated that this might be a common problem. It should be noted, however, that there is evidence to suggest differing levels of concordance between patients and proxies according to the dimension being assessed, which may introduce an element of bias into the PRO assessments.

Missing data may take two main forms: (1) item non-response, in which data are missing for one or more individual items, or (2) unit non-response, in which the whole questionnaire is missing for a participant, because of missing forms, the participant dropping out of the study or entering it late. Although it is possible to prevent much missing data, a certain amount should be expected within any study. The protocol should contain clear instructions on what procedures to follow in the event of a missing questionnaire or missing data for individual items, such as whether the participant should be contacted. In all cases, it is good practice to maintain a record of the reasons for missing data, in order to ascertain the extent to which this was related to the patient's PRO and to inform the analyses and interpretation of the data. The relative amount of missing data, the assumed cause, the sample size and the type of intended analyses will determine the degree to which missing data is a problem, and it will critically inform the interpretation of the results of the study.

Dealing with missing data

Poor compliance resulting in missing data can have a significant detrimental impact on the analyses and interpretation of data. First, fewer observations may result in the power of the study to detect an effect being compromised. Second, missing data may introduce selection bias into a trial and thereby compromise its validity, particularly if low compliance is associated with less well patients who have poor outcomes. It is therefore important that the impact of missing data is carefully addressed, and that the potential cause of the missing data is understood, as the most appropriate method for dealing with it will depend largely on the assumed mechanism by which it is missing: if the reasons for missing data are completely unrelated to the respondent's health, it is classified missing completely at random. If the likelihood of missing data depends only on previous scores but not current or future scores, it is considered missing at random. Data are only considered not missing at random if the 'missingness' is actually related to the value of the missing observation. The relevant approach for dealing with each type of missing data varies depending on whether potential bias due to data not missing at random needs to be addressed. Methods are available to test the assumption of whether or not data are missing at random. Having a record of the reasons for missing data in a trial is particularly important here.

Missing values for individual items may be imputed (filled in) to complete the data in order that a full analysis may be undertaken. Most commonly, the mean of answered questions is imputed, provided that at least half the items are completed, although this may not be suitable for scales whose items are ordered hierarchically according to difficulty. Other less common approaches include imputation based on the responses of other respondents or regression of missing items based on other scale items. Some instruments are provided with instructions on how to score the question when items are missing. Methods for dealing with missing whole questionnaires/assessments in longitudinal (repeated measures) studies when missing data are assumed to be ignorable include complete case analysis, which involves removing patients with incomplete data, and available case analysis, which uses all available cases on a specific variable and is subsequently more desirable. Alternatively, data can be filled in using imputation-based methods such as last observation carried forward, single mean imputation, predicted value imputation or hot-deck imputation. Statistical techniques such as likelihood-based methods have been developed for non-ignorable missing data, but they are complex and controversial and should therefore be applied with caution. More sophisticated models have been developed specifically for non-ignorable missing data, which are likely to arise in trials, but they are limited by their complexity and lack of availability and interpretability. They are often accompanied by sensitivity analyses, which are used to compare the appropriateness of employing a given strategy.

Analysing patient-reported outcomes and integrating results with clinical outcome data

The analyses of PRO data require appropriate methodological expertise, and collaboration with a statistician is recommended. If the research question and study design are clearly stated in the protocol and the PRO hypotheses established, then analysis of data should be possible and overexploring and overinterpreting data avoided. Using descriptive statistics to illustrate the impact of surgery on PROs is the initial first step and data may be presented as means with confidence intervals or in proportions (of patients with good or poor PRO scores). At present, there is still a lack of general understanding about the impact of most surgical treatments on PROs, and therefore using graphical methods to illustrate these simple descriptive changes will aid patients and surgeons.

One of the areas that further work is needed for surgeons to understand PROs within a clinical context is the meaning of changes in PRO scores over time. Because of unfamiliarity with questionnaires and scoring systems it can be difficult to use the data in practice. It is recommended that the use of validated tools and understanding the data themselves is more widespread to aid interpretation of the information and what changes in scores constitute a clinically significant difference.

It is also recommended that clinical outcomes are considered alongside PROs. For example, in a study reporting HRQL scores for cancer surgery, it is inappropriate to report that scores were very good a year after the operation without considering the survival of the group undergoing surgery. In studies of oesophagectomy for cancer, in which 2 year survival rates are approximately 50%, if those who survive have good HRQL then information to be imparted to patients before surgery needs to contain both accurate PRO and survival data for informed treatment choices to be made.

Conclusions

The use of PROs to evaluate surgery has only just begun to be established in healthcare and it is expected that over the next two decades the real value and place of this type of outcome measure will be better understood. In the meantime, it is recommended that more well-designed studies evaluating PROs of surgery are performed, analysed and reported, and that data are used routinely alongside clinical outcomes to inform clinical decision-making. Surgeons require training in the use of communicating PROs, and decision aids for each surgical procedure need developing to assist these processes. Accurate PRO data in the consultations will complement communication of short- and long-term clinical outcomes. The development of core outcome sets for surgical procedures including core PROs is one means of evaluating surgery in ways that will allow cross-centre, -surgeon and -study comparisons and data synthesis to occur. Future research into the prognostic value of PROs and their role in improving individual patient care and follow-up is still needed. It is anticipated, however, that better understanding of PROs will improve surgical practice and allow treatment decisions to be more patient centred and thus lead to better adherence to treatment and surgeon and patient satisfaction with this key part of surgical practice.

GUIDE TO FURTHER READING

Aaronson NK, Ahmedzai S, Bergman B, *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: a quality of life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;**85**:365–376.

Beauchamp T, Childress JF. *Principles of Biomedical Ethics*, 5th edn. Oxford, UK: Oxford University Press, 2001.

Bekker HL, Hewison J, Thornton JG. Understanding why decision aids work: linking process with outcome. *Patient Educ Couns* 2003;**50**:323–329.

Blazeby JM, Williams MH, Alderson D, Farndon JR. Observer variation in assessment of quality of life in patients with oesophageal cancer. *Br J Surg* 1995;**82**:1200–1203.

Brehaut JC, Fergusson DA, Kimmelman J, *et al.* Using decision aids may improve informed consent for research. *Contemp Clin Trials* 2010;**31**:218–220.

Cella DF, Tulsky DS, Gray G, *et al.* The Functional Assessment of Cancer Therapy Scale: development and validation of the general measure. *J Clin Oncol* 1993;**11**:570–579.

Collin C, Wade DT, Davies S, Horn V. The Barthel ADL Index. A reliability study. *Int Disabil Stud* 1988;**10**:61–63.

Efficace F, Bottomley A, Osoba D, *et al.* Beyond the development of health-related quality-of-life (HRQOL) measures: a checklist for evaluating HRQOL outcomes in cancer clinical trials -does HRQOL evaluation in prostate cancer research inform clinical decision making? *J Clin Oncol* 2003;**21**:3502–3511.

Elwyn G, Laitner S, Coulter A, *et al.* Implementing shared decision making in the NHS. *BMJ* 2011;**341**:971–975.

Euroqol Group. Euroqol – a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208.

Faden R, Beauchamp T, King N. *A History and Theory of Informed Consent*. New York, NY: Oxford University Press, 1986.

Fayers P, Machin D. *Quality of Life Assessment, Analysis and Interpretation*. Chichester, UK: Wiley, 2000.

Gilbody SM, House AO, Sheldon TA. Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ* 2001;**322**:406–409.

Gotay CC, Kawamoto CT, Bottomley A, Efficace F. The prognostic significance of patient-reported outcomes in cancer clinical trials. *J Clin Oncol* 2008;**26**:1355–1363.

Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Eval Clin Pract* 1999;**5**:401–416.

Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? *Soc Sci Med* 2005;**60**:833–843.

Hickey AM, Bury G, O'Boyle CA, *et al.* A new short form individual quality of life measure (SEIQoL-DW): application in a cohort of individuals with HIV/AIDS. *BMJ* 1996;**313**:29–33.

Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: Macleod CM (ed.) *Experimental Cancer Therapy*. New York, NY: Columbia University Press, 1949:191–205.

Katz S, Downs TD, Cash HR, Grotz RC. Progress in development of the index of ADL. *Gerontologist* 1970;**10**:20–30.

Kirkham JJ, Dwan KM, Altman DG, *et al.* The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;**340**:c365.

McCulloch P, Ward J, Tekkis PP. Mortality and morbidity in gastro-oesophageal cancer surgery: initial results of ASCOT multicentre prospective cohort study. *BMJ* 2003;**327**:756–761.

McNair AG, Brookes ST, Davis CR, *et al.* Communicating the results of randomized clinical trials: do patients understand multidimensional patient-reported outcomes? *J Clin Oncol* 2010;**28**:738–743.

Quinten C, Coens C, Mauer M, *et al.*; EORTC Clinical Groups. Baseline quality of life as a prognostic indicator of survival: a meta-analysis of individual patient data from EORTC clinical trials. *Lancet Oncol* 2009;**10**:865–871.

Royal College of Surgeons of England. Guidance on Surgical Practice: Criteria, Standards and Evidence. London, UK: RCS, 2004.

Ruta D, Garratt AM, Abdalla MI, *et al.* The SF 36 health survey questionnaire. *BMJ* 1993;**307**:448.

Ruta DA, Garratt AM, Leng M, *et al.* A new approach to the measurement of quality of life. The Patient-Generated Index. *Med Care* 1994;**32**:1109–1126.

Sanderson T, Morris M, Calnan M, *et al.* An additional core-set generated by people with rheumatoid arthritis: RA Patient Priorities for Pharmacological Interventions (RAPP-PI). *Arthritis Care Res* 2010;**62**:647–656.

Smets EM, Garssen B, Bonke B, de Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res* 1995;**39**:315–325.

Sprangers MAG, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J Clin Epidemiol* 1992;**45**:743–760.

Takeuchi EE, Keding A, Awad N, *et al.* Impact of patient-reported outcomes in oncology: a longitudinal analysis of patient-physician communication. *J Clin Oncol* 2011;**29**:2910–2917.

Tugwell P, Boers M, Brooks P, *et al.* OMERACT: an ongoing evidence based initiative to improve the quality of outcome measures in rheumatology. In: Tugwell P, Shea B, Boers M, *et al.* (eds) *Evidence Based Rheumatology*. London, UK: BMJ Publishing Group, 2004:19–24.

Tugwell P, Boers M, Brooks P, *et al.* OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials* 2007;**8**:38.

US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. *Guidance for Industry – Patient Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims*. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

Velikova G, Booth L, Smith AB, *et al.* Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 2004;**22**:714–724.

CHAPTER 6

Intensive and high-dependency care

ARUP CHAKRABORTY, NICOLAS PRICE AND MARTIN STOTZ

> *Salus aegroti suprema lex.*
>
> (Hippocrates c. 460 BC to c. 370 BC)

What is intensive care?

Intensive care medicine (ICM) is the science and art of identifying and managing critically ill patients while preventing further deterioration.

Intensive care demands a dedicated area in which all monitoring and therapeutic devices are immediately available to the appropriately trained professionals and a holistic approach of all professionals involved to diagnose, evaluate and integrate the patients' needs. This will help to develop objectives of care for the patients' best short-, medium- and long-term outcomes.

This chapter focuses on ICM and shall shed light on different aspects of this specialty. We will discuss advanced haemodynamic monitoring, vasopressor and inotropic therapy, sepsis, mechanical ventilation and, finally, death, which are all integral and important parts of modern ICM. It will not allow enough knowledge to make the careful reader a high-dependency or ICM specialist, but will allow an insight into some topics specifically related to the surgical patient. The aim of this chapter is to provide a fundamental understanding of underlying pathophysiology to facilitate a timely initiation of appropriate therapy by non-intensive care practitioners.

Evolution of a new specialty

Historically, Florence Nightingale was probably the founder of ICM. Nightingale had developed an outstanding dedication to patient care during the Crimean war (1853–6). She was named 'the lady with the lamp' as she was making rounds at night for the injured soldiers. As Florence recognized that some patients needed more frequent and careful monitoring than others, she started to group these patients closer to the nursing station. An area for high-dependency patients was born.

However, it was only in the last century, during the 1952 poliomyelitis epidemic, hospitals in Copenhagen, Denmark, started to create the spaces dedicated to perform mechanical ventilation for the treatment of multiple patients with respiratory failure. This new science, and the medical developments for specialized organ support that followed, needed the creation of designated areas in hospitals: the *Intensive Care Unit* (ICU). In parallel to these advances in technology, generations of physicians and nurses acquired new knowledge and skills over the years to form an integrated package of care, known today as *Intensive Care Medicine*.

Definition of intensive care

Intensive care can be broadly defined as a service for patients who have potentially recoverable conditions and who can benefit from more detailed observation and invasive treatment than that provided safely in an ordinary ward. It is usually reserved for patients who are at risk of, or indeed have, established organ failure. This can arise either as a result or complication of an acute illness or trauma, or as a predictable phase after major surgery.

The borders of intensive care and high-dependency care are fluid and not clearly defined. The progression of a disease over time will define the patient's need for more or less intense therapy. Areas in hospitals are usually classified by the level of care provided within, or more specifically by a nurse to patient ratio, which matches the level of care that can be provided (Table 6.1). The high-dependency unit (HDU) usually consists

Table 6.1 Example of levels of care

Levels of care	Nurse to patient ratio	Patient status
Level 1	4:1	At risk of deterioration Relocation from higher levels of care
Level 2	2–2.5:1	More detailed observation or intervention necessary, including support of a single organ system 'Step down' from higher levels of care
Level 3	1:1	Advanced monitoring and support of >2 organ systems Advanced respiratory support

of level 2 care, whereas level 3 care is reserved for the ICU. This concept has been well tested and is an approved model of meeting medical demands.

It is the concentration of the skills, expertise and resources coming together in one designated area that makes an ICU. However, HDU and ICU care should never be seen as isolated specialties and co-operation with other specialists is mandatory to achieve the best possible patient outcomes.

'Open' and 'closed' Intensive Care Unit models

Today, different models of ICUs exist across the world. Full-time, fully trained intensive care physicians run most ICUs in Europe and Australasia. The intensivist is leading a multidisciplinary team of experts who are able to provide cover for specialized ICU care for 24 hours per day, 7 days per week; this is also known as a *closed* ICU model. This model has evolved since the 1960s, taking into account the development and definition of this unique specialty. The model of care in the USA, however, is different, with a significant number of ICUs still using the *open* model. This means that the decisions about care and therapy of patients remains with the primary physician, who is not necessarily a member of the ICU permanent staff but who leads the patient's management from hospital admission to discharge. The primary physician will, in turn, look after patients on the ICU, with the input of an intensive care professional only if needed.

As the closed model has been shown to decrease both ICU and hospital mortality and length of stay for patients, it seems favourable compared with the open model. Designated staff and around-the-clock availability can best provide continuity of care on the ICU/HDU. It also underlines the fact that ICM has evolved into a separate specialty that requires appropriate training to achieve best outcomes. Hence, ICM has become a specialty in its own right and is best provided by specifically trained professionals.

Criteria for admission to the intensive care unit

Nowadays, ICU professionals seem to become more involved in the care of unstable patients outside the ICU, a movement called 'critical care without walls'. ICM is expensive and labour intensive, thus resources in hospitals may be limited. A significant responsibility is on the intensivist together with the primary team to appropriately identify and triage those patients for ICU and HDU admission who are most likely to benefit from ICM and to restrict admission of others. ICU admission criteria must be able to cope not only with the critically ill emergency patient but also with planned elective admissions after major surgery.

The development of admission criteria is a complex issue and generally depends on multiple factors:

- the medical status of the patient at referral
- any underlying medical conditions, including dependency of the individual before admission
- ethical and religious implications for the patient (including the wishes of the patient, if known) and sometimes needs of the wider society (e.g. epidemic outbrakes such as H1N1)
- availability of level 2/3 care; if this is not available on site, the availability of a transfer team must be considered.

Attempts have been made to define admission (and discharge) policies based on scientific grounds, and were first published by the Society of Intensive Care Medicine in the USA. In Europe, several regional and national guidelines exist, but a conclusive pan-European consensus is missing to date. Most intensivists will usually base the decision whether to admit patients on a combination of prioritizing criteria, such as those outlined in Table 6.2.

Identifying critical illness

Care of the acutely ill patient in hospital may require input from critical care. The ageing population and increasing complexity of medical and surgical care mean that patients in hospital are at increasing risk of becoming acutely ill and may require admission to ICU areas. However, recognition of acute deterioration of the unwell can be delayed or even missed, as outlined in several reports.

The minimum standard should record the following physiological observations:

- heart rate (HR)
- respiratory rate
- systolic blood pressure
- level of consciousness
- oxygen saturation
- temperature.

Table 6.2 Priorities for admission to an intensive care unit

Priority	Patient	Intervention needed	Limitations of treatment	Special considerations
1	Critically ill Unstable physiology Emergency or planned admission	Advanced monitoring Intensive care most likely	None	Brain death confirmed or expected (consider organ donation)
2	Critically ill Moderately unstable Emergency or planned admission	Advanced monitoring Possibly intensive care	None	
3	Critically ill Instability of various degrees	Advanced monitoring Possibly intensive care	Perhaps	Underlying disease Reduced likelihood of recovery
4	Instability of various degrees		Yes or no	'Too well to benefit' 'Too sick to benefit' Refusal by patient

Physiological track-and-trigger systems are useful to monitor all patients in acute hospital settings. Staff should have competencies in monitoring, interpretation and response to the acute deterioration. If abnormal physiology is detected, a graded response should be triggered. This should be agreed and delivered locally. If admission to a critical care area is indicated, the decision to admit the patient should involve both a senior decision-maker of the primary team caring for the patient on the ward and the ICU.

Patients can be classified into three groups and should trigger responses as follows:

- Low score group:
  - increased frequency of observations and informing the senior nurse.
- Medium score group:
  - urgent call to the primary medical team
  - simultaneous call to personnel with competencies in acute illness.
- High score group:
  - emergency call to the team with critical care competencies. These competencies should include assessment of the critically ill patient, advanced airway management and resuscitation skills.

Several physiological track-and-trigger systems are in use. Single parameter systems are simple to use and are easy to reproduce. Multiple parameter systems are more complex but allow for better monitoring of clinical progress and for a graded response strategy, but may lack reproducibility.

Whatever track-and-trigger system is used locally, escalation policies should be agreed as well as follow-up of the recognized deterioration. To support this, some institutions use specialized response teams or critical care outreach that can be contacted if a patient deteriorates.

Haemodynamic monitoring

Although clinical examination can assist guiding fluid and inotrope therapy, it is not always accurate. Advanced haemodynamic monitoring has established itself as an invaluable tool to guide therapy in the critically ill patient. It encompasses examination and measurement of the arterial pressure waveform, the central venous pressure (CVP) and cardiac output (CO). And, indeed, haemodynamic monitoring has been shown to improve outcome in a variety of settings. This section will discuss some elements of advanced haemodynamic monitoring and the underlying cardiovascular physiology.

Physiology

To a certain extent, CO and organ perfusion can be assessed clinically with a combination of the following:

- feeling the temperature of the extremities
- assessing the capillary refill time
- assessing the conscious level
- measuring the urine output
- blood pressure measurement.

In addition, blood lactate and base deficit can also give an indirect indication of global organ perfusion. However, clinical assessment of CO has a great interobserver variation and different observers may give widely differing opinions of the CO and filling status of the same patient. One reason for the use of more scientific methodology to quantify CO is to attain a degree of standardization and reproducibility.

Cardiac physiology

CO is the product of heart rate (HR) and stroke volume (SV), i.e. CO = HR × SV, whereas SV is determined by the cardiac preload, afterload, and contractility.

The cardiac index (CI) is obtained by dividing the CO by the patient's body surface area {BSA; BSA (m^2) = $\sqrt{[\text{height (cm)} \times \text{weight (kg)}]/3600}$} allowing comparison of individuals by adjusting for height and weight:

$$\mathrm{CI} = \mathrm{CO}/\mathrm{BSA}$$

Blood pressure = CO × vascular resistance; therefore systemic vascular resistance (SVR) and pulmonary vascular resistance (PVR) are:

$$\mathrm{SVR} = [(\mathrm{MAP} - \mathrm{CVP}) \times 80]/\mathrm{CO}$$

$$\mathrm{PVR} = [(\mathrm{PAP} - \mathrm{PAWP}) \times 80]/\mathrm{CO}$$

where MAP is the mean arterial pressure, CVP is the central venous pressure, PAP is the pulmonary arterial pressure and PAWP is the pulmonary arterial wedge pressure (see Fig. 6.4). The factor of 80 in the above equations serves for unit conversion (dynes s/cm^5).

Physiology of oxygen

Blood oxygen content (CaO_2) is the sum of the oxygen bound to haemoglobin (Hb-O_2) and the oxygen dissolved in plasma: 1 g of haemoglobin is able to bind a maximum of 1.34 mL of oxygen, whereas 1 g of plasma binds 0.003 mL of oxygen:

$$CaO_2 = (\mathrm{Hb\text{-}O_2}) + (PO_2 \times 0.003)$$

$$\mathrm{CaO_2} = ((\text{Hb oxygen saturation}/100) \times \mathrm{Hb} \times 1.34) + (PO_2 \times 0.003)$$

Delivery of oxygen (DO_2) is the product of CaO_2 and CO:

$$DO_2 = CaO_2 \times \mathrm{CO}$$

$$DO_2 = ((\text{arterial oxygen saturation} \times \mathrm{Hb} \times 1.34) + (PO_2 \times 0.003)) \times \mathrm{CO}$$

Perhaps the most important goal in the management of the critically ill patient is the maintenance of adequate DO_2. The amount of oxygen dissolved in plasma is so small that, for most purposes, it can be almost ignored. O_2 consumption (VO_2) is obtained by multiplying the CO by the difference in arterial and mixed venous oxygen content:

$$VO_2 = \mathrm{Hb} \times 1.34 \times ((\text{arterial saturations}/100) - (\text{mixed venous saturations}/100)) \times \mathrm{CO}$$

Table 6.3 Summary of equations with normal ranges

Variable	Calculation	Normal range
CI	CO/BSA	2.5–4.2 L/min/m^2
SI	SV/BSA	40–60 mL/m^2
SVR	((MAP – RAP) × 80)/CO	960–1400 dyn.s/cm^3
PVR	((PAP – PAWP) × 80)/CO	25–125 dyn.s/cm^3
DO_2	CO × ((saturations/100) × Hb × 1.34)	950–1300 mL/min
VO_2	CO × Hb × 1.34 × ((arterial saturations/100) – (mixed venous saturations/100))	180–320 mL/min

BSA, body surface area; CI, cardiac index; CO, cardiac output; DO_2, oxygen delivery; MAP, mean arterial pressure; PAP, pulmonary arterial pressure; PAWP, pulmonary arterial wedge pressure; PVR, pulmonary vascular resistance; RAP, right atrial pressure; SI, stroke index; SV, stroke volume; SVR, systemic vascular resistance; VO_2, O_2 consumption.

Some of the equations with normal ranges are summarized in Table 6.3.

Following this concept, Shoemaker and peers aimed to improve survival in critically ill patients through therapeutically driving DO_2 and VO_2 to supranormal levels by early aggressive fluid therapy and by using vasopressors and inotropes in the early 1990s. After initial promising results, other authors were not able to confirm this hypothesis and the concept has been abandoned these days. Moreover, evidence has emerged recently that, rather than a decrease in overall oxygen delivery, microcirculatory abnormalities and an inability to extract oxygen (dysoxia) seem to prevail in critical illness.

Arterial pressure

Non-invasive measurement of blood pressure is intermittent. Continuous invasive arterial pressure monitoring is often employed in ICUs as real-time measurement of blood pressure and is mandatory for patients treated with vasoactive medication. In addition, frequent arterial blood gas analyses are often necessary in critically ill patients who receive mechanical ventilation.

The radial, femoral and brachial arteries are the most commonly used sites for insertion. The arterial pressure is transduced via a short extension of rigid manometer tubing. The transducer is zeroed to atmospheric pressure and should be positioned at the level of the patient's heart (or axilla).

A marked respiratory swing of the arterial wave can signify hypovolaemia, especially in a ventilated patient. Hypovolaemia can also lead to a low dicrotic notch (Figure 6.1). The SV can be derived from the arterial waveform by analysis of the area under the systolic portion (see section "Measuring cardiac output").

Central venous pressure

The CVP is most frequently measured using the internal jugular and subclavian veins. While the femoral vein can be used as well, it is noteworthy that values of femoral CVP are less accurate than those of the internal jugular and subclavian veins. The central venous catheter is connected to a transducer system that is identical to that of arterial catheters.

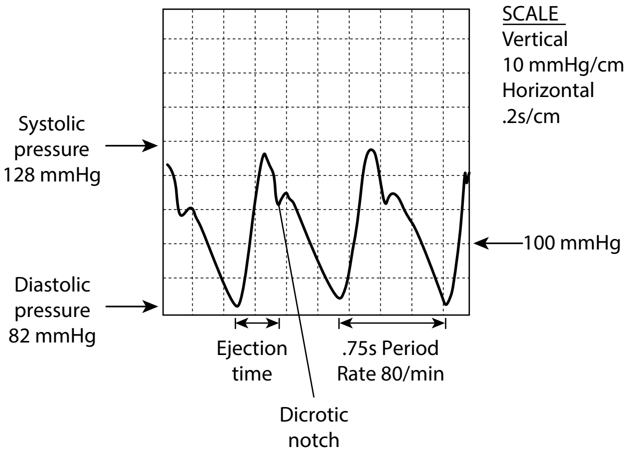


Figure 6.1 A typical arterial wave. Note that the dicrotic notch in a peripheral arterial wave is thought to be the pressure wave reflected peripherally rather than the closure of the aortic valve. Scale: vertical, 10 mmHg/cm; horizontal, 2 s/cm.

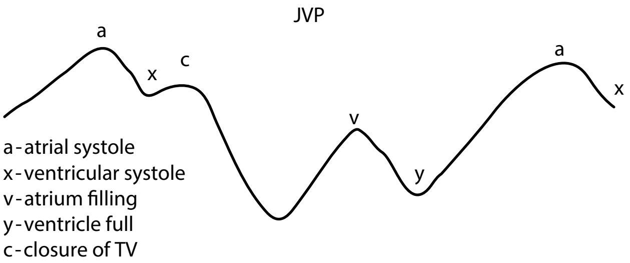


Figure 6.2 Components of central venous pressure (CVP). a, atrial systole; x, ventricular systole; v, atrium filling; y, ventricle full; c, closure of tricuspid valve.

In spontaneously breathing patients the CVP should be measured at end-expiration as it decreases during spontaneous inspiration (by a suction effect of the lungs), whereas in patients undergoing positive-pressure ventilation measurements should be taken at end-inspiration. A typical CVP waveform is displayed in Figure 6.2.

Determinants of CVP are:

- blood volume
- venous tone
- right heart compliance
- intrapleural pressure
- patient position.

See also section "Fluid responsiveness".

Central and mixed venous oxygen saturation

Central venous oxygen saturation can be measured from blood taken from the superior vena cava. In turn, the mixed venous oxygen saturation is acquired by taking blood from the pulmonary artery. In theory, central and mixed venous oxygen saturation differ only in the addition of the venous return of the heart muscle to the latter, but in practice values do not always correlate (Table 6.4). Targeting the central venous oxygen saturation is an important part of early goal-directed therapy in sepsis (see section "Sepsis").

Table 6.4 Mixed venous and central venous oxygen saturation explained

Low venous oxygen saturation (high peripheral oxygen extraction)	**High venous oxygen saturation (low peripheral oxygen extraction)**
Increased O_2 consumption or metabolic rate, e.g. sepsis (early stage)	Decreased O_2 consumption, e.g. hypothermia
Low cardiac output, e.g. cardiogenic shock	Increased cardiac output, e.g. pregnancy, vasodilatation
Hypoxaemia	Left-to-right shunt
Anaemia	Sepsis (after adequate resuscitation)

Measuring cardiac output

We know invasive and non-invasive methods of measuring the CO, as described below.

Invasive

These include:

- the Fick method
- dye dilution
- thermodilution
- pulse contour analysis.

Non-invasive

These include:

- doppler
- thoracic bioimpedance
- echocardiography.

Fick method

The Fick method utilizes the principle that the amount of oxygen consumed is equal to the difference in oxygen concentration between the mixed venous and arterial systems multiplied by the CO. VO_2 is measured in breath samples from a reservoir bag using a spirometer. A pulmonary artery catheter is used to obtain mixed venous blood whereas an arterial line is used to determine the other variable in the Fick equation:

$$CO = VO_2/(\text{arterial} - \text{mixed venous } O_2 \text{ content difference})$$

This method is rarely used in clinical practice.

Dye dilution

The dilution of a defined volume of dye or indicator into an unknown volume of fluid and measurement of its concentration over time can indirectly measure the CO. Indocynanine green can be used but a substance much more commonly used in clinical practice is lithium (LiDCO system). Dye can be injected into either a peripheral or a central vein and sampled from a normal arterial catheter and CO can be derived. To maintain accuracy intermittent calibration with injection of small, non-toxic doses of lithium is suggested. LiDCO is becoming increasingly popular as it is less invasive than pulmonary artery catheterization and interobserver variation is not a problem as with other techniques for advanced haemodynamic monitoring such as oesophageal Doppler.

Thermodilution

This technique is utilized by both the pulmonary artery catheter (PAC) and the PiCCO system to measure CO. The PAC (Figure 6.3) has an inflatable balloon at the tip that allows floating of the PAC with the bloodstream into the pulmonary artery. This is guided by the pressure waveforms at different points (right atrium, right ventricle, pulmonary artery and wedge position).

The proximal lumen, 30 cm from the tip, lies in the right atrium and measures CVP. The thermistor lumen is situated 4 cm from the tip of the catheter and measures temperature. The distal lumen is at the tip of the catheter, lies in a branch of the pulmonary artery, and is connected to a pressure transducer. The balloon lumen permits introduction of 1.5 mL of air into the balloon at the distal tip.

The pressure waveforms at different anatomical points are illustrated in Figure 6.4. These are used to determine the position of the pulmonary catheter. As the PAC is floated, it passes from the right atrium to the right ventricle, the systolic pressure increases and the diastolic pressure stays the same.

When advanced further, the PAC passes from the right ventricle to the pulmonary artery, typically indicated by an increase in diastolic pressure. As the PAC passes from the pulmonary artery into the occlusion (or wedge) position, the pressures decrease (to less than the pulmonary artery diastolic pressure).

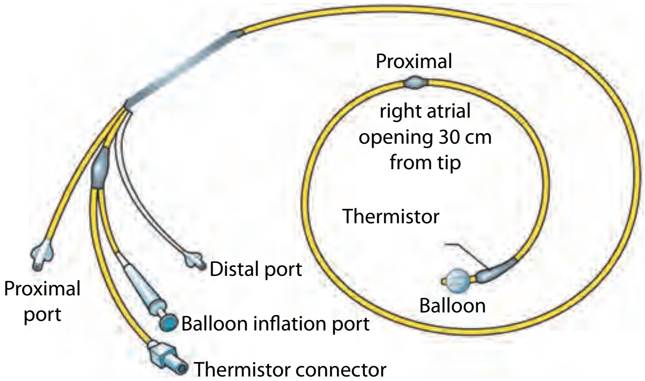


Figure 6.3 Pulmonary artery catheter.

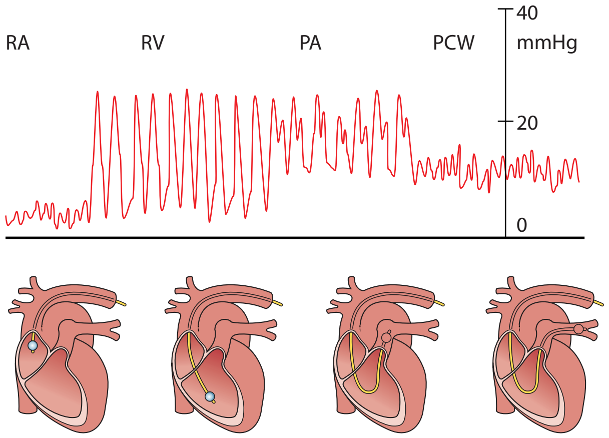


Figure 6.4 Change of pressure wave form (top) according to the location (bottom) of the tip of the Swann–Ganz catheter. PA, pulmonary artery; PCW, pulmonary capillary wedge pressure; RA, right atrium; RV, right ventricle.

The PAC can be used to directly measure or derive the following:

- the CO: a 10 mL bolus of cold fluid is injected through the proximal lumen into the central circulation (right atrium) and the temperature change of the circulating blood is detected by the distal thermistor near the catheter tip. CO is calculated using the drop of temperature over time as volume and temperature of the injected solution are known.
- RAP: measured directly.
- right ventricular pressure: measured directly.
- pulmonary artery pressure: measured directly.
- left atrial pressure: this correlates with the occlusion pressure. This is the pressure measured by occluding ('wedging') a pulmonary catheter with an inflated balloon. The tip distal to the balloon measures the pressure in a continuous column of blood that extends from the catheter tip, through the pulmonary capillaries and veins, and into the left atrium. The pulmonary artery occlusion pressure ('wedge') pressure can give an indication of the filling status of the left ventricle.
- mixed venous oxygen saturation: this is obtained by taking blood from the distal lumen.

Indications for pulmonary arterial catheterization
These include:

- left and right heart pressures do not correlate:
  - right heart failure
  - left heart failure
  - pulmonary hypertension
  - valvular heart disease.
- fluid status difficult to assess:
  - heart failure
  - renal failure.
- left ventricular ejection fraction unknown/echocardiography unavailable.
- guide use of vasoactive drugs, fluids and diuretics.

Contraindications of pulmonary arterial catheterization
These include:

- coagulopathy
- tricuspid valve pathology
- pulmonary valve pathology.

Insertion of a PAC is not without complications. Beside the well-known risks of central venous cannulation these specifically include arrhythmias, pulmonary artery rupture and lung infarction, as well as damage to the heart valves and catheter knotting.

As with CVP readings, wedge pressure readings should be taken at end-expiration (the highest part of the wedge waveform), whereas for ventilated patients pressures should be taken at end-inspiration.

Several studies have put the use of the PAC for management of critically ill patients in question. A prospective cohort study from 1996 has suggested increased mortality of patients managed with PAC, whereas other studies including *PACman* in 2005 and *ARDSnet* in 2006 failed to show an overall benefit of the use of PAC. Not surprisingly, use of the PAC on ICUs has declined significantly in recent years as other less invasive and less complex methods of advanced haemodynamic monitoring have become available to the medical practitioners. However, PAC is still the technique of choice for measurement of pulmonary artery pressures.

Pulse contour analysis

SV can be derived from the area under the systolic portion of the arterial waveform. To increase accuracy, waveform analysis to calculate CO can be calibrated against a standard curve generated with either dye dilution (LiDCO) or thermodilution (PiCCO). Unlike LiDCO, a special arterial catheter with a thermistor and a central venous catheter are required for the PiCCO system. Both systems automatically calibrate to derive CO and other parameters from continuous pulse contour analysis. Besides CO, these devices also claim to be able to determine other parameters that are thought to be helpful in patients' management such as stroke volume variation (SVV), pulse pressure variation (PPV) and extravascular lung water as well as providing an estimate for cardiac contractility. SVV and PPV are considered measures of preload and fluid responsiveness, respectively, which will be described later.

Other devices use algorithms, unique to each manufacturer, and claim to derive CO from waveform analysis without calibration. Examples of pure pulse contour analysis devices are Vigileo and LiDCOrapid.

Oesophageal Doppler

The Doppler effect is a change in the observed frequency of a sound wave occurring when the source and observer are in motion relative to each other, with the frequency increasing when the source and observer approach each other and decreasing when they move apart. Oesophageal Doppler monitoring (ODM) is perhaps the most common form of CO monitoring in use in UK ICUs and operating theatres today. A probe is placed in the oesophagus via the oral or nasal route. The probe generates a low-frequency ultrasound signal, which is reflected by red blood cells moving down the descending aorta. By applying the Doppler principle, the reflected signal is proportionate to the flow velocity. A normogram is used to account for the patient's weight, age and sex as the area of flow for measurement is equally important. SV and CO can be derived from blood flow time; peak velocity is an indicator for myocardial contractility. Fluid management is guided by the use of ODM by observing changes in SV following a fluid challenge. This approach can correct functional hypovolaemia and optimize intravascular volume. This approach has shown promising results to the extent that the National Institute for Health and Clinical Excellence (NICE) in the UK has recently recommended the use of ODM in patients undergoing major or high-risk surgery.

Thoracic bioimpedance

Bioimpedance is defined as the electrical resistance of tissue to the flow of current. When small electrical signals are transmitted through the thorax, the current travels along the blood-filled aorta, which is the most conductive area. Changes in bioimpedance, measured at each beat of the heart, are inversely related to pulsatile changes in volume and velocity of blood in the aorta. CO thus can be calculated from bioimpedance.

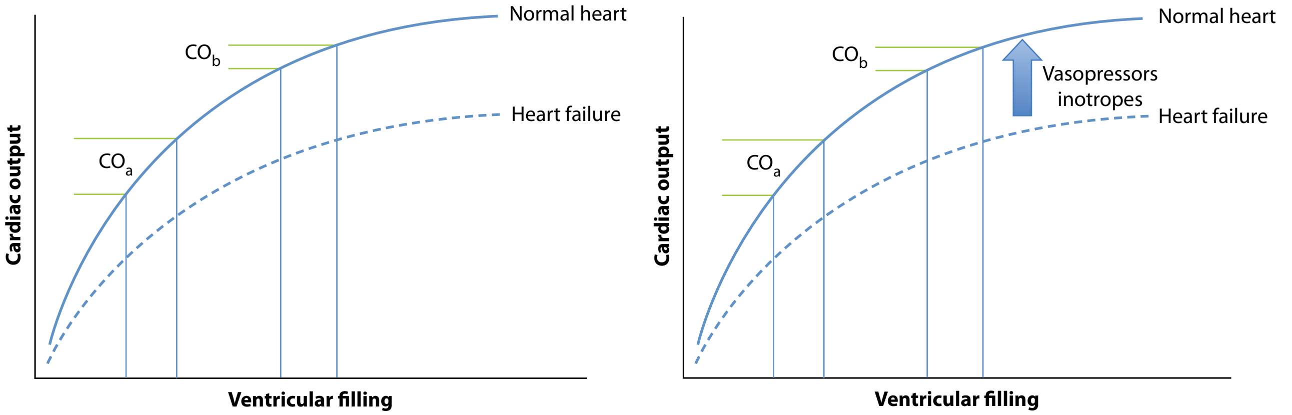


Figure 6.5 The Frank–Starling relationship shows the dependency of cardiac output (CO) on left ventricular filling. A defined change of left ventricular filling leads to a higher increase in CO on the steep slope of the curve (ΔCO_a) than on the flat part of the curve (ΔCO_b). Heart failure changes this relationship (dashed line); changes in ventricular filling need to be carried out more carefully to avoid ventricular overload.

Echocardiography

The use of echocardiography is increasing in operating theatres, emergency medicine and ICUs. Echocardiography is non-invasive and can be used at the bedside at all times. In addition, it gives results in real time and can be repeated for perioperative monitoring and in critically ill patients. The main disadvantages are that it needs an operator trained in its use and that the method is highly operator dependent. Furthermore, obtaining images in intubated patients can be technically challenging.

Transoesophageal echocardiography (TOE) uses higher ultrasound frequencies than transthoracic echocardiography (TTE) to produce better resolution images. TOE is less available than TTE but might be more useful in the ICU setting.

Goal-directed therapy

Optimizing central venous oxygen saturation, SV or CO has been shown to improve outcome. Improved outcome has been shown for the following:

- severe sepsis
- pre- and perioperatively for colon surgery, hip fractures and other major or high-risk surgery
- postoperatively, including cardiac surgery and major or high-risk surgery.

Typically, fluid management is guided by observing changes in SV following fluid challenges to achieve optimal CO. The aim of this approach is to correct functional hypovolaemia and optimize intravascular volume according to the Frank–Starling relationship (Figure 6.5).

The evidence shows that optimization of cardiac performance reduces complications after surgery and length of hospital stay, and decreases the use of central venous catheters, therefore improving systems' performance and reducing healthcare costs. Although the evidence for use of goal-directed fluid therapy in the ICU settings is not as clear, there is still widespread use.

Sepsis

Sepsis, severe sepsis and septic shock represent increasingly grave stages of the systemic inflammatory response to severe infection (Figure 6.6). Reported mortality rates of severe sepsis and septic shock still remain at 40–60% today and have not changed significantly in the past 20 years. Despite intense research and recent advances in treatment, patients with sepsis suffer considerable long-term morbidity. This section tries to elucidate the underlying cellular mechanisms of sepsis. The progression from a simple infection to septic shock and multiorgan failure is discussed and current treatment strategies explained.

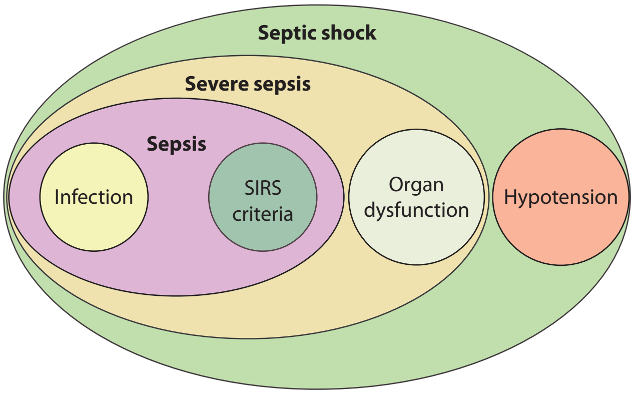


Figure 6.6 The different severities of sepsis may be a continuum. SIRS, Systemic inflammatory response syndrome.

Infection and immunity

A considerable number of bacteria exist on our internal and external surfaces. Mucosa of skin and gut is the first barrier and line of defence to prevent bacterial invasion of the human host. However, when this barrier fails or is disrupted (i.e. by surgical wounds, indwelling catheters, burns), the host becomes susceptible to pathogen invasion. The second-line defence is the immune system that recognizes, fights and destroys invading germs. Sepsis represents systemic response of the immune system to infection. Both pathogen virulence and host resistance regulate the severity of the inflammatory response. The resistance and immune competence of the host are determined by multiple factors such as age, sex, genetic predisposition, nutritional status and underlying health conditions. The inborn or innate immune system consists of cellular and humoral components, both orchestrating the host immune response.

The cellular and humoral components of the innate immune system

The cellular component of the innate immune system includes circulating monocytes, tissue macrophages, neutrophils and lymphocytes. Tissue macrophages are capable of engulfing and digesting microbes and will recruit and attract other phagocytes by secreting cytokines. Macrophages present particles of dispatched microbes as antigens to lymphocytes and hence interact closely with the adaptive immune system. Neutrophils are important phagocytes, but can also destroy invading pathogens by secreting toxic substances such as reactive oxygen species (ROS). Neutrophils attracted by chemokines migrate and translocate into the infected tissue to fight pathogens. More neutrophils and immature forms are liberated from the stimulated bone marrow, leading to a final increased number in blood (neutrophilia and leftward shift of the neutrophils). Eosinophil and basophil granulocytes are responsible for secreting inflammatory mediators and creating an inflammatory environment. These lead to dilation of tight junctions of the adjacent vessels, facilitating the migration of more inflammatory cells into the infected tissue and leading to efflux of plasma. As a consequence of these processes, signs of local inflammation (rubor, calor, dolor) occur. Of note, cells of the innate immune system can fight invading pathogens directly without involvement of the adaptive immune system.

Cytokines are the humoral components of the innate immune system and act either directly on pathogens or as mediators between cells and organs. ROS in high concentrations have a toxic effect on bacteria by damaging their cell walls. ROS in lower concentrations seem to act as important regulatory mediators. Interleukins (IL) and tumour necrosis factor (TNF) are other mediators that have the ability to promote inflammation from a localized to a systemic process. Some of these cytokines are stored in myeloid cells and can be secreted rapidly after contact with pathogens.

Pathogen recognition and activation of the inflammatory cascade

Cells of the innate immune system can detect typical molecular patterns of most microbes, including viruses, bacteria, fungi and protozoa. Examples of such patterns are lipopolysaccharides (endotoxin) from the cell wall of Gram-negative bacteria; lipoteichoic acid and peptidoglycan from Gram-positive bacteria; unmethylated bacterial DNA; or double-stranded RNA from viruses. These molecules are recognized by three families of specific pattern recognition receptors: toll-like receptors (TLRs), intracellular nucleotide-binding proteins, and peptidoglycan recognition proteins. To date, 10 different types of TLRs have been identified in humans. Most act in conjunction with other molecules such as CD14, or with other TLRs expressed on the cell surface. Binding of a microbial molecule to its specific TLR results in signal transmission by the toll/IL-1 receptor homologous region adaptor proteins to a complex intracellular cascade of enzymes. These enzymes consist of kinases – enzymes that phosphorylate and thus activate proteins necessary for nuclear gene transcription and production of cytokines. Although the most well-known, nuclear factor κB (NF-κB) is only one example of many transcription factors activated during sepsis. One of the genes expressed via NF-κB codes for TNF, which further amplifies local NF-κB activation (Figure 6.7). TNF is then released in the systemic circulation to reach effector sites. On a systemic level, TNF and IL-6 induce the production of acute phase proteins in the liver, e.g. C-reactive protein, procalcitonin and fibrinogen. TNF also plays an important role in the activation of programmed cell death or apoptosis.

Another activated enzyme of interest is inducible nitric oxide synthase (iNOS). In sepsis, iNOS is upregulated and generates high levels of nitric oxide (NO), a fundamental proinflammatory molecule. NO activates other enzymes such as guanylyl cyclase, which leads to the production of cyclic guanosine monophosphate (cGMP). cGMP and NO are responsible for local and systemic vasodilatation, which can lead to hypotension and shock. Newer data seem to suggest that NO may have beneficial effects that are determinants in patients' survival.

Anti-inflammatory activity during sepsis

To prevent an overwhelming and possibly deleterious proinflammatory response anti-inflammatory mechanisms are activated simultaneously during infection. These include secretion of specific cytokines such as IL-10 and the soluble TNF receptor leading to a decrease in lymphocyte cell count. The overall balance is presumed to be proinflammatory in the early phase of sepsis and anti-inflammatory later. The advantage of this systemic anti-inflammatory response may be the attenuation of deleterious

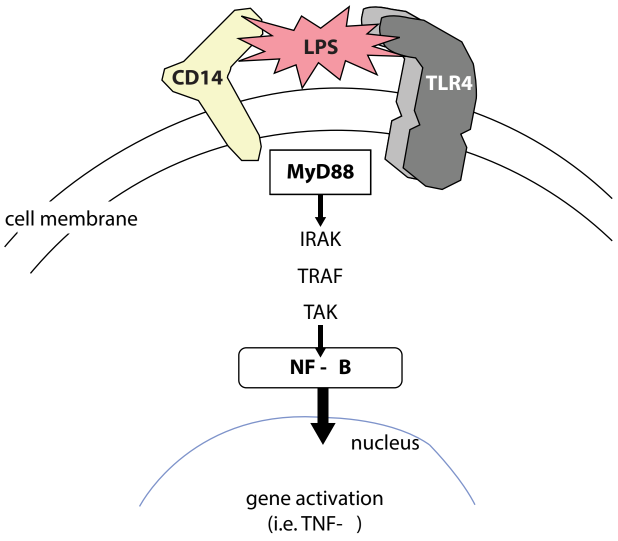


Figure 6.7 Activation of a CD14+ T-helper cell by a fragment of Gram-negative bacteria. IRAK, interleukin-1 receptor-associated kinase; LPS, lipopolysaccharide (endotoxin); NF-kB, nuclear factor kB; TAK, TRAF-associated kinase; TLR, toll-like receptor; TNF, tumour necrosis factor; TRAF, tumour necrosis factor receptor-associated factor. LPS from the Gram-negative bacterial wall binds to TLR4 and CD14, activating the toll/IL-1 receptor homologous region domain called myeloid differentiation protein 88 (MyD88). This activates the IRAK, which stimulates the TRAF and, consequently, the TAK. As a result, the nuclear transcription factor NF-kB is liberated from its inhibitor and is able to dislocate into the cell nucleus to bind to DNA, activating hundreds of specific genes coding for proteins. (With permission of EMH Swiss Medical Publishers Ltd; www.smw.ch.)

systemic proinflammatory effects and the concentration of the inflammatory response at the site of infection. However, when anti-inflammatory mechanisms dominate, the immune system is depressed (immunoparesis), thus increasing the body's susceptibility to nosocomial infections and to dormant pathogens such as cytomegalovirus. Whether patients with severe sepsis might benefit from immune system stimulation is the topic of current investigations.

Microcirculation and mitochondrial dysfunction

Early sepsis is characterized by vasodilatation and intravascular volume depletion (from increased capillary leak and external losses), leading to a low CO. This is made worse by myocardial depression and microvascular distribution alterations and causes an oxygen supply–demand imbalance in various organ beds. Fluid administration during early sepsis increases oxygen delivery to the organs and is known to improve outcome (see below). Within the cells, mitochondria require oxygen to produce ATP by oxidative phosphorylation. ATP is the predominant source of intracellular energy and is needed for all energy-dependent cellular functions. Mitochondria will suffer structural and functional damage by the inflammatory insult following infection by three separate mechanisms: (1) direct inhibition of the respiratory enzyme complexes by increased concentrations of nitric oxide and its metabolite peroxynitrite, (2) increased production of ROS and (3) genetic downregulation of new mitochondrial protein formation. This damage is facilitated by overwhelmed mitochondrial defence mechanisms (reduced glutathione, superoxide dismutase). Loss of mitochondrial function is associated with electron microscopic structural changes such as membrane thickening, swelling and vacuolization of mitochondria and finally loss of mitochondrial integrity and death.

Mitochondrial biogenesis

Failure on a mitochondrial level of energetic status in patients with sepsis-induced multiorgan failure has been shown recently. However, repair mechanisms exist and mitochondria may be repaired or regenerated by mitochondrial biogenesis. Mitochondrial biogenesis is triggered by increased expression of transcription factors in the nucleus. This will result in restoration of the energetic status and ATP production over time and promote survival of critical illness. This may prove to be a useful therapeutic target for accelerating organ recovery, and stimulation of mitochondrial biogenesis (such as with NO) may play an important role in the future.

Hibernation

Macro- and microcirculatory failure will cause tissue hypoxia by insufficient oxygen supply, whereas impaired oxygen utilization owing to mitochondrial dysfunction will lead to tissue dysoxia. Both mechanisms decrease intracellular ATP production. This not only may impair organ-specific cell function but may also result in a loss of energy-dependent cell integrity. Hence, lack of ATP will lead to cellular dysfunction and, eventually, cell death. However, postmortem studies of septic patients failed to reveal micro- and macroscopic evidence of cell death.

This lack of necrosis has led to the hypothesis that organ dysfunction may represent a functional rather than a structural phenomenon and thus may be potentially reversible. Therefore, cells may shift the use of ATP only to processes essential for cell survival. By shutting down energy-dependent organ-specific functions the cell can decrease its total ATP expenditure, allowing the net ATP balance to remain positive despite a decrease in ATP production. This 'suspended animation' is analogous to aestivation and hibernation. Although this is a new concept for multiorgan dysfunction in sepsis, it is a well-established protective strategy in a variety of animals as an adaptive response to heat, cold or drought. Hence, organ dysfunction during sepsis can be seen as an adaptive and even protective process that will help to prevent cell death. Once infection is overcome and the organism returns back to 'normal', mitochondrial function is restored. ATP production then increases and the cell can regain its normal metabolism and organ-specific function.

Definitions and diagnosis of sepsis

Despite being a frequent reason for admission to the ICU, infections and sepsis can be difficult to diagnose. The definitions in Table 6.5 may be helpful and provide the clinician with the appropriate nomenclature for the diagnosis and management of infections. Of note, systemic inflammatory response syndrome (SIRS), sepsis, severe sepsis and septic shock are thought to represent increasing severity of disease (Figure 6.7).

Different stresses can stimulate the human body to liberate cytokines, which can clinically present as a SIRS. Therefore, SIRS may be seen not only following infection but also after major surgery and severe trauma, making the diagnosis of sepsis potentially difficult. However, the search for the origin of the SIRS is of the utmost importance, as this time delay may prove significant for the patient.

A few features of the definitions of sepsis, however, warrant a closer look. First, the definition of sepsis needs only SIRS and the suspicion of infection (positive cultures are *not* essential for the diagnosis of sepsis). This warrants a thorough review of the patient at the bedside by an experienced clinician. Second,

Table 6.5 Nomenclature for the diagnosis and management of infections

Bacteraemia	**Positive blood cultures**
Infection	Invasion of the body by organisms
SIRS	Two or more of the following: Temperature >38°C or <36°C Heart rate >90/min Respiratory rate >20/min or minute volume >10 L/min in ventilated patients or $PaCO_2$ <4.3 kPa White cell count >12 000/mL or >10% immature forms or wcc < 4 × 10^2/L
Sepsis (former sepsis syndrome)	SIRS *plus* the clinical suspicion of infection
Severe sepsis	Sepsis *plus* evidence of organ dysfunction
Septic shock	Sepsis *plus* hypotension refractory to volume resuscitation
Multiorgan failure	Failure of >2 organ systems

SIRS, systemic inflammatory response syndrome.

the diagnosis of septic shock includes a therapeutic intervention such as fluid resuscitation. This is an important feature as the failure to respond to fluid resuscitation is the discriminating factor between severe sepsis and septic shock.

But, rather than seeing sepsis as an isolated disease, it is now the belief that sepsis represents a disease *complex* with potential involvement of several organ systems (multiorgan failure). As described above, sepsis often presents clinically with hypoperfusion of organs.

The clinical signs of decreased organ perfusion include:

- low blood pressure and hypotension
- oliguria
- confusion/drowsiness
- angina or signs of myocardial ischaemia on 12-lead ECG
- metabolic (usually lactic) acidosis.

Most common locations of the septic focus are:

- lungs
- abdomen
- genitourinary tract
- skin
- indwelling catheters
- central nervous system.

Therapeutic principles: the 'golden hour' of sepsis

The principles of sepsis treatment include early goal-directed therapy, source identification and control, early antibiotics, steroids, activated protein C and tight glycaemic control (Figure 6.8).

Early goal-directed therapy

Early adequate fluid resuscitation and source control of the infectious focus are still the cornerstones of therapy for sepsis. To minimize delay in treatment, ideally diagnosis and therapy should be rolled out simultaneously. In trauma care, the 'golden hour' after the insult is a widely recognized concept that should be equally applied to the treatment of sepsis.

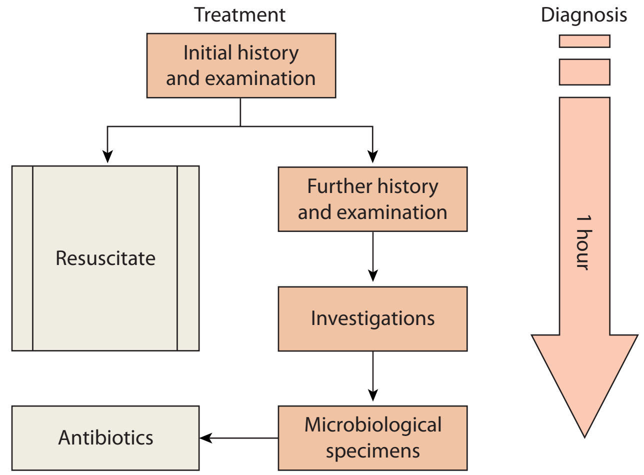

Treatment
Initial history and examination
Diagnosis
Resuscitate
Further history and examination
1 hour
Investigations
Antibiotics
Microbiological specimens

Figure 6.8 The 'golden hour' of sepsis. Diagnosis and treatment need to be performed simultaneously to minimize time delay.

The goal of fluid resuscitation is to achieve and maintain adequate end-organ perfusion (early goal-directed therapy). This will minimize mitochondrial damage and finally organ failure. Rivers and co-workers were able to show a significant decrease in hospital mortality due to sepsis from 46.5% in the standard group to 30.5% in the early goal-directed therapy group by introducing an early (within 6 hours of admission) and aggressive therapeutic approach. Whether crystalloids or colloids are the optimal choice for resuscitation is still under debate (see Shock). Fluid challenges according to the patient's needs and to optimize CO are more appropriate than the blind administration of maintenance fluids.

In an initial phase, 250 mL of colloid or 500 mL of crystalloid should be administered rapidly (i.e. 10–15 minutes) to the septic patient and titrated against clinical response and reversal of clinical signs of hypoperfusion (restoration of skin perfusion, blood pressure, urine output, metabolic acidosis). The insertion of a central venous line, urinary catheter and arterial line may help facilitate diagnosis and treatment in severe cases. It is of special importance not to pass the threshold of tolerance of the cardiovascular system and induce congestion. Rivers and his peers advocated fluid administration for a CVP of 8–12 mmHg. If blood pressure is not restored with adequate fluid challenges alone, vasopressors such as norepinephrine (noradrenaline) or vasopressin might be considered to target a MAP of >65 mmHg (see Vasopressors and inotropes). Dopamine is still used widely as a vasopressor in sepsis, but has less favourable 28 day mortality than norepinephrine. A MAP target of >65 mmHg has never been tested in RCTs and is only founded on expert opinion, but to target a higher MAP and expose patients to higher doses and prolonged treatment with vasopressors seems to have less favourable outcomes.

The next goal of early goal-directed therapy is to restore adequate CO. Having achieved restoration of fluid status and blood pressure, aiming for a central venous saturation ($ScvO_2$) of >70% has been advocated. $ScvO_2$ is a used as a surrogate for peripheral oxygen extraction assuming that peripheral oxygen extraction is inversely related to CO (see Haemodynamic monitoring). To achieve $ScvO_2$ >70%, first a haematocrit above 30% (haemoglobin >10 g/L) should be targeted using allogenic blood transfusion if necessary. If this is achieved, inotropes such as dobutamine or epinephrine (adrenaline) may be used to increase CO. The use of a combination of norepinephrine and dobutamine for septic shock has comparable outcomes to the use of epinephrine alone.

Source identification and control

Simultaneously to the initiation of treatment of sepsis, identification and eradication of the source of infection is crucial. If time allows, appropriate samples for microbiology should be taken and additional imaging should be considered before initiation of antibiotic therapy. Sampling for microbiology processing should include blood cultures taken from all indwelling catheters and simultaneously taken peripheral stabs, respiratory samples as well as sample fluids from all drains (urinary catheter, wound drains, spinal or ventricular drains). Sputum is preferred in the non-intubated patient, whereas directed or non-directed bronchial lavage provides more accurate results in patients with artificial airways. Gram staining may help to facilitate the choice of treatment in this 'urgent choice' antibiotic strategy, but has its pitfalls.

If additional imaging reveals a septic focus (e.g. intra-abdominal abscess, obstructive cholecystitis), image-guided percutaneous drainage may be considered at the same time. This offers diagnostic as well as therapeutic options, which may avoid the patient having to undergo more surgery.

Early antibiotics

In contrast to clinical situations in which the start of antibiotics can be delayed, patients with severe sepsis and septic shock must receive appropriate antibiotics within the first hour of diagnosis. Each hour of delay in antibiotic administration for the initial 6 hours is associated with a decrease in survival.

The choice of antibiotic is dependent on several factors, such as the patient's underlying disease, whether the infection is community or hospital acquired and the local antimicrobial resistance. Immunocompromised or neutropenic patients, including those after splenectomy, are likely to have acquired different infections from immunocompetent patients, and the choice of antimicrobial treatment must account for this special situation. In contrast to community-acquired infections, hospital-acquired infections tend to involve more resistant pathogens, or at least pathogens with a different antimicrobial spectrum, which, again, needs to be considered.

It is beyond the remit of this chapter to give specific recommendations for antibiotic treatment, but the principle of treatment for severe sepsis and septic shock should always entail early *broad-spectrum* antimicrobial cover, following the previous considerations, and narrowing of antimicrobial spectrum after receiving microbiology results. Guidelines for antimicrobial treatment are especially helpful for the use of antibiotics by the less experienced and have been shown to be beneficial for patients' outcome. Guidelines should also consider the duration of treatment for common clinical situations as prolonged antibiotic treatment can promote microbial resistance. To avoid increasing local antibiotic resistance, regular input from microbiologists and infection control teams should be sought. ICUs are areas where antibiotic consumption and resistance rates are highest. This gives the intensivist the special responsibility in antibiotic stewardship to avoid spiralling antimicrobial resistance.

Steroids in sepsis

For decades the use of corticosteroids in sepsis has been and is still controversially discussed. Besides the common and well-known effects of corticosteroids, advantageous specific effects in sepsis include inhibition of cytokines and iNOS (see above) and potential to reverse septic shock. Disadvantages include hyperglycaemia, risk of superinfections and an increased incidence of critical illness polyneuromyopathy.

Since the late 1980s several trials have failed to show survival benefit from the use of pharmacological doses of corticosteroids in sepsis (i.e. hydrocortisone 100 mg t.d.s.). In addition, timing of treatment (early versus late) did not show beneficial effects on outcome. However, this research also revealed that sepsis has the potential to increase adrenal dysfunction and tissue resistance to corticosteroids. This new insight has allowed a follow-up on the concept of steroid treament. Finally, Annane and co-workers were able to demonstrate beneficial effects of low-dose early treatment with corticosteroids (hydrocortisone 50 mg q.d.s. and fludrocortisone 50 μg o.d.) in patients with inadequate adrenal response. Based on this multicentre RCT, many physicians worldwide have started to use steroids for patients in septic shock requiring vasopressors.

Just a few years later, another multicentre RCT of low-dose early corticosteroids failed to confirm these beneficial effects. However, this so-called *Corticus trial* was also able to confirm earlier reversal of shock with the use of hydrocortisone. This has finally led the Surviving Sepsis Campaign to recommend that 'hydrocortisone should be considered in the management strategy of patients with septic shock after blood pressure is poorly responsive to fluid resuscitation and vasopressor therapy'.

Tight glycaemic control

Stress-induced hyperglycaemia is the elevation of blood glucose in the presence of acute illness. Stress hormones, endogenous and exogenous, such as epinephrine and cortisol as well as cytokines, are contributing factors, all potentially inhibiting insulin release and action, therefore enhancing gluconeogenesis and inhibiting glycogen formation. In turn, administration of intravenous glucose, parenteral nutrition and a series of other medications can also contribute to hyperglycaemia. On one hand, hyperglycaemia can reflect the severity of disease; on the other, hyperglycaemia can be linked with worse outcomes in non-diabetic patients admitted to the ICU as duration of exposure to higher glucose concentrations is inversely associated with survival.

Only since the turn of the century has evidence for glucose control on the ICU emerged and tight glycaemic control (TGC) been advocated for the treatment of critically ill patients. Van der Berghe and co-workers were able to demonstrate more favourable outcomes in a large single-centre trial if blood glucose was kept in a tight range (4.4–6.1 mmol/L) than with a more conventional approach (target 10–11 mmol/L). Patients on TGC had decreased mortality, fewer days of ventilation and shorter length of stay on the ICU than controls; however, significantly more episodes of hypoglycaemia were reported in the treatment group. As this trial was carried out in a selected group of patients, mostly after cardiac surgery, a subsequent trial including general medical patients failed to show the same beneficial effect. In contrast, it has also become evident that hypoglycaemia is associated independently with increased mortality in the critically ill. Therefore, it is the recent recommendation to target a blood glucose range of 6–10 mmol/L in non-diabetic critically ill patients. This is achieved best by administering short-acting insulin by continuous intravenous infusion according to local protocols; measures to minimize and treat hypoglycaemia are important. Evidence for the perioperative period is less conclusive.

Vasopressors and inotropes

Shock

Shock is a state of impaired tissue perfusion insufficient to meet the metabolic demands of tissues. Clinical signs of impaired perfusion can still be subtle, but, owing to tissue hypoperfusion, increased lactate production from a switch to anaerobic metabolism at the cellular level may already be present. Global tissue perfusion depends upon blood flow, and MAP (product of CO and SVR) is often used as a clinical surrogate. It is important to note that, although a low blood pressure is synonymous with shock, it does not strictly speaking form part of this definition. This is because during shock a compensatory tachycardia and/or vasoconstriction can result in preservation of a relatively normal blood pressure. Shock can be categorized by the underlying pathophysiological mechanism:

- hypovolaemic
- cardiogenic
- obstructive
- distributive.

Hypovolaemic shock can result from dehydration, haemorrhage or burns and is usually associated with a compensatory vasoconstriction and increase in SVR to maintain vital organ perfusion. Cardiogenic shock is most often a result of myocardial ischaemia but can be seen in association with arrhythmias, valvular heart disease or after cardiac surgery. It can be biventricular or affect predominantly either the left or right side of the heart. In cardiogenic shock, the CI is <2.2 L/min/m^2 (Table 6.3) despite an adequate ventricular preload. Cardiogenic shock due to critical myocardial ischaemia has a high mortality of approximately 50%, which can be significantly reduced with urgent revascularization. Obstructive shock is caused by mechanical obstruction of blood flow, e.g. by cardiac tamponade or massive pulmonary embolism. Management should be directed at the cause of the obstruction, such as pericardiocentesis for cardiac tamponade. Distributive shock is manifest by a markedly reduced vascular tone, mostly from the release of vasodilatory inflammatory mediators. There is often a compensatory increase in the CO. This is the commonest type of shock found in sepsis, but is also seen with high spinal cord injuries with peripheral vasodilatation from loss of sympathetic tone and anaphylactic reactions.

In clinical practice, these distinctions are often arbitrary, and more than one form of shock can coexist. Septic patients are usually (relatively) hypovolaemic, may have evidence of left ventricular hypokinesia and have distributive shock (see section "Sepsis").

Fluid therapy

First-line management for all types of shock involves the administration of intravenous fluids as well as treating the underlying cause. In cardiogenic shock, fluid resuscitation needs to be applied with caution to avoid left ventricular failure resulting in pulmonary oedema. Often, the preload can be optimized by administering fluid boluses carefully, particularly when right ventricular impairment is present (e.g. with an inferior myocardial infarct). The relationship between left ventricular end-diastolic volume (a surrogate for preload) and CO was described by Frank and Starling (Figure 6.5); the therapeutic process aims to shift the CO to the peak of the curve, but note that, for the failing myocardium, the slope of the initial part of the curve is more gradual. However, assessing the optimal fluid administration is far from easy and a more useful concept of fluid responsiveness has been described.

Fluid responsiveness

Accurate clinical assessment of fluid status is notoriously difficult, but thorough clinical examination is essential to look for signs of hypovolaemia such as dry mucous membranes and reduced capillary refill. Measurement of CVP is often helpful, especially the haemodynamic response to a fluid challenge/higher CVP. However, it is important to note that the pressure measured depends upon the compliance of the vena cava, which is in turn dependent upon vasomotor tone. There is poor correlation between the CVP and left ventricular end-diastolic volume, and hence the ability of the CVP to accurately inform the clinician of the likely improvement in CO with fluids is unreliable. A high CVP can, however, be helpful as it suggests adequate preload and hence a low likelihood of a successful response to a fluid bolus. A suitable challenge is 250 mL of fluid given over a short period of time (e.g. 15 minutes) with a comparison of physiological variables such as CO or MAP before and after the challenge. CO monitoring can be very helpful for this purpose. Up to 50% of patients with septic shock will show no improvement in CO when given fluid loading. Once patients no longer respond to fluid resuscitation, therapeutic considerations should include using inotropes and vasopressors.

Inotropes/vasopressors

Inotropes act primarily by increasing the intracellular calcium concentration (via increased cyclic AMP) available for excitation–contraction coupling in the cardiac myocyte, and hence increasing myocardial contractility and finally improving CO. Vasopressors, on the other hand, increase vascular tone by activation of sympathetic-mediated vasoconstriction (Table 6.6), which is mostly receptor mediated. Most commonly used agents have mixed actions and work on more than one receptor (Table 6.7). Most vasopressors have short half-lives and need to be given by continuous intravenous infusion; potent vasoconstrictors must be given by a central venous line to avoid peripheral complications.

Table 6.6 Classification of cardiovascular adrenergic receptors

Receptor type	Site	Action
α_1	Vascular smooth muscle	Vasoconstriction
β_1	Heart	Increase in heart rate and contractility
β_2	Vascular smooth muscle	Relaxation

Table 6.7 Commonly used vasopressors and inotropes

Drug	Dose range	Receptors				Notes
		a	β_1	β_2	Dopaminergic	
Epinephrine	0.01–0.1 µg/kg/min	+++++	++++	+++		Can cause raised serum lactate, hyperglycaemia, metabolic acidosis and arrhythmias
Norepinephrine	0.01–3 µg/kg/min	+++++	+++	++		Can cause decreased cardiac output if used in hypovolaemic patients Probably the commonest vasopressor used in the ICU
Dopamine	0–5 µg/kg/min				++	Effects vary according to infusion rate
	5–10 µg/kg/min	+	+	+	++	Low-dose dopamine does not reduce the incidence of renal failure
	10–20 µg/kg/min		+	+	++	
Dobutamine	2–20 µg/kg/min	+	+++++	+++		Adverse effects include tachycardia, arrhythmias, hypotension (from β_2-mediated vasodilatation)
Vasopressin	0.01–0.1 U/min					Acts on V_1, V_2 and V_3 receptors State of relative vasopressin deficiency can develop in prolonged shock
Milrinone	Bolus 50 µg/kg over 10–30 minutes Infusion 0.375–0.75 µg/kg/min					PDEIII inhibitor. Used for severe congestive cardiac failure Contraindicated with left ventricular outflow tract obstruction (e.g. HOCM, aortic stenosis)
Levosimendan	Loading dose 12–24 µg/kg over 10 minutes Infusion 0.05–0.2 µg/kg/min					Binds to troponin C, causing positive inotropy, and opens ATP-sensitive K^+ channels, causing vasodilatation Used to improve left or right ventricular function after cardiac surgery and short-term haemodynamic improvement in acute heart failure

HOCM, hypertrophic obstructive cardiomyopathy; ICU, intensive care unit; PDEIII, phosphodiesterase III.

It is crucial to ensure adequate fluid resuscitation before considering inotropes or vasopressors. Shock and subsequent activation of the sympathetic nervous system may cause selective perfusion of vital organs and may exert changes in regional blood flow. This will happen at the expense of perfusion of the gut and skin, resulting in relative ischaemia, release of endotoxins and, hence, worsening of the inflammatory cascade. This mechanism can be exacerbated by vasopressors. If inotropes are used in hypovolaemic patients, this has the potential to increase the (already elevated) myocardial oxygen consumption, which, if unmatched by oxygen supply, will cause myocardial ischaemia.

Norepinephrine (noradrenaline)

Norepinephrine is an endogenous catecholamine (it is the primary neurotransmitter at postganglionic sympathetic neurones) with a predominantly α_1 agonist activity and minor β_1 agonist activity (especially at low doses). It causes arteriolar and venous vasoconstriction and hence increases peripheral resistance with little change in CO.

Epinephrine (adrenaline)

This naturally occurring catecholamine acts as a potent agonist at α and β receptors. Its effects also vary according to the dose; at low doses, β effects predominate, but with higher doses α-mediated vasoconstriction becomes more apparent. It is most commonly used in patients with a low CO unresponsive to other inotropes. Epinephrine often causes a tachycardia and can reduce splanchnic blood supply as well as increase serum lactate, which is not necessarily related to organ hypoperfusion. It is used most often for anaphylaxis, cardiac arrest and in patients with severe bronchospasm.

Dopamine

Dopamine is the natural precursor to norepinephrine and epinephrine. It acts on dopamine receptors as well as α and β receptors. Its effects vary according to the infusion rate:

- 0–5 µg/kg/min: renal and splanchnic vasodilatation (via dopamine receptors)
- 5–10 µg/kg/min: increase in HR and contractility (β_1 effect)
- 10–20 µg/kg/min: vasoconstriction (α_1 effect).

Although it does increase urine output, low-dose dopamine does not reduce the risk of renal failure in the critically ill. Dopamine is more arrhythmogenic and increases HR and SV more than norepinephrine.

Dobutamine

This synthetic catecholamine is primarily a β_1 agonist, but also has mild β_2 and weak α_1 activity. It causes an increase in HR and SV, thereby increasing CO. It is often given in combination with norepinephrine to counteract the β_2 effects of vasodilatation. It will increase HR and myocardial oxygen consumption, which may be deleterious if coronary blood flow is impaired.

Vasopressin

Vasopressin is an endogenous peptide released from the posterior pituitary in response to hypotension/hypovolaemia or increased serum osmolality. It causes vasoconstriction via stimulation of

V_1 receptors on vascular smooth muscle cells. In septic shock there is often a relative vasopressin deficiency; however, there is little convincing evidence to suggest it should be used as a first-line vasopressor, hence it is normally used in conjunction with norepinephrine or dopamine. Terlipressin, which has a longer half-life, is sometimes used as an alternative.

Other agents

The phosphodiesterase inhibitors *milrinone* and *enoximone* are sometimes used, particularly in patients with heart failure or after cardiac surgery. These agents inhibit cyclic AMP breakdown by inhibition of phosphodiesterase III, thus causing positive inotropy as well as facilitating diastolic relaxation (lusitropy) and cardiac filling. Phosphodiesterase inhibitors are usually administered as an intravenous bolus followed by a continuous infusion. They are both potent vasodilators owing to their β_2 effect, and often patients need treatment with a vasopressor simultaneously to maintain MAP.

Levosimendan

This is an inodilator used for acutely decompensated chronic heart failure. It binds to troponin C, increasing myocardial calcium sensitivity, and has a positive inotropic effect. It is also a potent vasodilator via opening of ATP-sensitive K^+ channels; hence, myocardial contractility is increased without increased myocardial oxygen consumption. As with the phosphodiesterase inhibitors, a vasopressor such as norepinephrine often needs to be infused as well to offset its vasodilatory effect.

Ventilation support

Ventilation support is a cornerstone of ICU and HDU care. HDUs and ICUs should be able to offer ventilation treatment and the medical team are familiar with this technique. As ventilation support does warrant specialized monitoring, it is obvious that this must be provided simultaneously. Some specialized services offer ventilation assist as well (e.g. weaning units, spinal units), but this exceeds the aim of this section.

Respiratory failure

In health pulmonary ventilation matches the demands of the body and will regulate O_2 uptake and CO_2 elimination. Pulmonary gas exchange is a combination of alveolar ventilation and lung perfusion. The normal alveolar ventilation–perfusion relationship (V/Q) in health is approximately 0.8 and any change in either ventilation or perfusion can alter this equilibrium and lead to respiratory failure. Respiratory failure can be acute or chronic. For more details also refer to Chapter 9 "Pulmonary insufficiency".

Two types of respiratory failure can be distinguished: whereas type 1 acute respiratory failure is defined by the inability to meet the patient's oxygen demand resulting in hypoxaemia, type 2 respiratory failure denotes the inability of CO_2 clearance and will typically result in hypercapnia. As the amount of alveolar CO_2 can accumulate to critical levels, severe hypercapnic failure will lead to hypoxia. In contrast, hypoxic failure may lead to respiratory fatigue, and accumulation of CO_2 can finally occur. A combination of both hypoxia and hypercapnia is therefore common. The alveolar–arterial (A–a) oxygen gradient can be of help to diagnose the cause of hypoxia.

Arterial blood gas analysis is the method of choice to diagnose disturbances of pulmonary gas exchange. Peripheral oxygen saturation and end-tidal CO_2 ($etCO_2$) as well as the respiratory rate and respiratory mechanics can point to an impairment of gas exchange but are not able to replace direct arterial blood gas measurement.

The equation for the A–a oxygen gradient is:

$$\text{A–a gradient} = PAO_2 - PaO_2$$

where PAO_2 is the alveolar PO_2 (calculated from the alveolar gas equation) and PaO_2 is the arterial PO_2 (measured in arterial blood).

The alveolar gas equation is (see Table 6.8 for explanation of the abbreviations and sample values):

$$PAO_2 = FIO_2(P_{ATM} - P_{H2O}) - (PaCO_2(1 - FIO_2(1 - RQ)))/RQ$$

Hypoxaemic, or type 1, respiratory failure is a potentially dangerous condition as cellular metabolism is dependent on oxygen. The human body can tolerate hypoxaemia for a certain period of time, especially when conditioned by high altitude or chronic pulmonary disease. The Extreme Everest Project has demonstrated recently that trained healthy volunteers were able to tolerate low blood oxygen tension (mean arterial oxygen partial pressure 3.28 kPa).

If oxygen administration via a nasal cannula or facemask (with rebreathing bag) does not improve the patient's hypoxic condition, breathing assist with continuous positive airway pressure (CPAP) should be considered. Hypercapnia is mainly a result of decreased alveolar ventilation and the lack of CO_2 clearance. Type 2 respiratory failure therefore warrants an increase in alveolar ventilation.

Assisted breathing

The goal of assisted breathing is to maintain continuous positive airway pressure (CPAP). This can only be achieved by the use of a tight fitting facemask. During spontaneous breathing, negative intrathoracic pressures will prevail during a normal breathing cycle. Therefore, assisted breathing and mechanical ventilation

Table 6.8 Explanation of the abbreviations and sample values for the alveolar gas equation

Quantity	Description	Sample value
P_AO_2	Alveolar partial pressure of oxygen (PO_2)	107 mmHg (14.2 kPa)
FIO_2	Fraction of inspired gas that is oxygen (expressed as a decimal)	0.21 (= 21%)
P_{ATM}	Atmospheric pressure	760 mmHg (101 kPa)
P_{H2O}	Saturated vapour pressure of water at body temperature and at atmospheric pressure	47 mmHg (6.25 kPa)
P_aCO_2	Arterial partial pressure of carbon dioxide (PCO_2)	36–45 mmHg (4.79 kPa)
RQ	The respiratory quotient	0.8

should not be mistaken for spontaneous breathing, even if the patient breathes spontaneously!

Positive intrathoracic pressures will impede venous return to the heart and may compromise CO. In susceptible individuals, this effect needs to be counteracted by additional fluid administration and/or haemodynamic support with vasopressors and inotropes. On the other hand, this effect can be used to decrease left ventricular preload in patients with acute heart failure.

During assisted breathing via facemask, inflation of the stomach may occur. A nasogastric tube on continuous drainage may be needed to prevent overinflation and to minimize the risk of pulmonary aspiration of gastric content. Enteral feeding while on continuous or intermittent assisted breathing therapy warrants careful consideration. Both breathing assist and the underlying pathology can cause distress to the patient, resulting in agitation and a lack of co-operation. If sedation and analgesia are considered, this must be carefully titrated to the desired effect to avoid respiratory and haemodynamic compromise.

PEEP, NIV or what? The 'babel' of assisted breathing

Several devices and techniques are available for ventilation assist. To complicate matters for the inexperienced, a uniform nomenclature does not exist. Hence, it may be useful to understand the underlying principle of assisted breathing.

For a start, positive end-expiratory pressure (PEEP) and continuous positive airway pressure (CPAP) are synonyms. To increase the alveolar partial pressure and to decrease the A–a gradient, oxygen can be delivered with positive pressure. Many devices will serve that purpose and are available on the market. A valve in the expiratory circuit is used to regulate the end-expiratory pressure (overflow). PEEP levels of 5–10 cmH_2O are usually safe to administer via facemask. If higher levels of PEEP are required to maintain sufficient oxygenation, endotracheal intubation may be considered.

PEEP will not only prevent end-expiratory alveolar collapse but has also the potential to reopen and recruit collapsed alveoli, resulting in an increase in the overall area of pulmonary gas exchange. PEEP, however, does provide only minimal assistance during inspiration and will lead to only a minimal increase in alveolar ventilation. PEEP is therefore predominantly of use in hypoxic failure.

For treatment of hypercapnia, an increase in either tidal volume or respiratory rate is needed to facilitate alveolar CO_2 removal. To achieve an increase in tidal volume, inspiratory volume support will be needed in addition to PEEP. The increase in tidal volume is facilitated by pressure support during inspiration (or flow, which in turn results in increased inspiratory pressures), as pressure and volume are directly correlated. Combined with PEEP, this will lead to a biphasic positive-pressure curve, or bilevel positive airway pressure (BiPAP) (Figure 6.9). Confusingly, BiPAP and non-invasive ventilation (NIV) are often used as synonyms in daily practice. In addition, NIV is used for the description of all ventilation assist that is not delivered through an endotracheal tube. For the purpose of this chapter, NIV is used as a synonym for BiPAP.

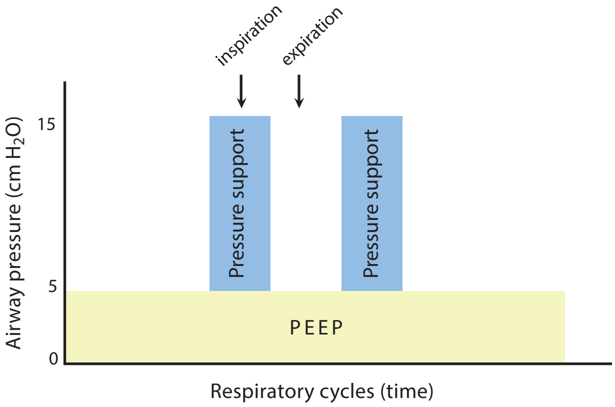


Figure 6.9 Illustration of PEEP and pressure support (NIV) in regard to airway pressures during the respiratory cycle.

NIV will assist to increase alveolar ventilation, resulting in a decreased alveolar CO_2. Many ventilation disorders are associated with additional work of breathing to compensate for the underlying condition. Pressure support and NIV have the potential to significantly reduce the work of breathing, therefore avoiding further patient treatment escalation.

Indeed, in two independent investigations, NIV therapy has been shown to prevent endotracheal intubation in up to 60% of patients with acute respiratory failure and finally to improve outcomes. Moreover, prophylactic use of CPAP and NIV postoperatively after major abdominal and lung resection surgery has been shown to reduce the reintubation rate by almost 40%.

Mechanical ventilation

Mechanical ventilation is usually delivered via an endotracheal tube (ETT). However, most NIV machines are able to deliver full mechanical and mandatory ventilation via facemask. Mechanical ventilation warrants a higher level of support and its use is limited to the ICU and the ICU specialized medical team. Patients on mechanical ventilation usually need more sedation and analgesia than for CPAP/NIV and are at high risk of suffering complications from overdosing and accumulation of analgosedation.

The most common modes of mechanical ventilation are found in Table 6.9. In contrast to short-term ventilation in the perioperative setting, mechanical ventilation on the ICU should allow more interaction of the patient (spontaneous breathing), to prevent and rebuild respiratory muscle loss.

Mechanical ventilation is not without risks, and evidence is accumulating that instrumentation of the upper airway with ETT or tracheostomy can be harmful. Recognized postintubation pulmonary complications can be divided into long term and short term. If excluding the immediate life-threatening complication of intubation, short-term complications consist predominantly of ventilator-associated pneumonia (VAP), whereas long-term complications include:

- airway symptoms (persistent hoarseness, stridor)
- tracheal stenosis
- cosmetic deformities.

Table 6.9 The most common modes in mechanical ventilation

Mode		Spontaneous breathing	Mandatory ventilation
IPPV	Intermittent positive pressure ventilation	–	+
CMV	Continuous mandatory ventilation	+	+
SIMV	Spontaneous intermittent mandatory ventilation	+	+
PCV	Pressure control ventilation	+	+
BiPAP	Bi-level positive airway pressure ventilation	+	+
PSV	Pressure support ventilation	+	–

The debate whether tracheostomy is superior to ETT for ventilation exceeding 5 days is currently ongoing. Also, whether early tracheostomy is more beneficial than late for patients' outcome remains unclear.

Ventilator-associated pneumonia

Postintubation pulmonary complications, and more specifically ventilator-associated pneumonia (VAP), have become well-recognized complications of mechanical ventilation. It is now accepted that long-term ventilation increases patients' morbidity and mortality *per se*. VAP does increase ICU and hospital stay and has significant cost implications, with increased antibiotic use and need for additional ICU resources and for the healthcare system.

VAP is defined as an airway infection after mechanical ventilation for more than 48 hours. VAP is presumed to originate from microaspiration of pharyngeal secretions along the cuff of the airway device (ETT or tracheostomy). The patient's underlying condition and pre-existing pulmonary pathology is a recognized cofactor. VAP is more commonly reported after certain types of surgical procedures, such as thoracic and cardiac surgery as well as complex vascular procedures. It is, however, difficult to prove causality arising from the surgical procedure itself, as many patients undergoing this type of surgery will have pre-existing pulmonary pathologies. Not surprisingly, duration of the operation seems to be as important as the time of postoperative mechanical ventilation.

Several care bundles have been proposed to prevent VAP in mechanically ventilated patients. These include regular mouthwashes for oral decontamination, establishing early enteral feeding and stress ulcer prophylaxis. Whether selective decontamination of the digestive tract, as practised in some ICUs, is able to prevent VAP is still under debate. New airway devices with changed cuff shapes and subglottic aspiration facility have shown promising preliminary results to decrease the incidence of VAP.

In conclusion, mechanical ventilation can cause VAP and cannot be seen as a low-risk procedure as it has an inherent morbidity and mortality.

Acute lung injury/acute respiratory distress syndrome

Severe pulmonary failure is defined by diffusion hypoxia due to a global decreased diffusion capacity of the alveolar membrane. Because of the lower ability of oxygen to cross membranes than CO_2, the increased diffusion distance will affect pulmonary oxygen gas exchange first. Traditionally, acute lung injury (ALI) and acute respiratory distress syndrome (ARDS) have been defined as a continuum of deteriorating gas exchange in the absence of left ventricular heart failure. Primary ALI/ARDS is defined by an underlying pathology directly affecting the lungs (i.e. pneumonia), whereas secondary ALI/ARDS is a consequence of pulmonary reaction to a systemic disease (i.e. massive transfusion). ALI/ARDS is caused by the loss of integrity of the alveolar membrane, which will lead to a thickening of the membrane and finally exudation of plasma into the alveoli as a result of capillary leakage. Histology has shown that ALI/ARDS will not uniformly affect the lung, but healthy and diseased alveoli will coexist in very close proximity. ALI/ARDS is a serious disease complex and has a reported mortality of up to 50% (see Chapter 9 "Pulmonary insufficiency" for more details).

The ratio of the arterial oxygen tension (PaO_2) to the fraction of inspired oxygen (FiO_2) is used to distinguish between ALI and ARDS (P/F ratio):

- ALI = P/F ratio <400 mmHg (53.3 kPa)
- ARDS = P/F ratio <200 mmHg (26.7 kPa).

Treatment of ALI/ARDS should be restricted to the ICU. Prevention of further harm while gaining time to heal is the therapeutic principle of mechanical ventilation for ALI/ARDS. Limiting tidal volumes to 6 mL/kg of ideal body weight and the utilization of high PEEP (up to 20 cmH_2O) has been shown to be beneficial to patients by the *ARDS Network*. Ventilation with low airway pressures is presumed to limit overdistension of the more rigid alveoli while minimizing harm to non-affected lung parenchyma. Intravenous fluid administration should be optimized to the minimum necessary to maximize CO. Changing of the patient's position with rotation beds and mattresses (even proning) will help to alter V/Q gravitational changes and should be considered in ALI/ARDS treatment. The concept of permissive hypercapnia, aiming to induce a left shift of the oxygen dissociation curve to facilitate peripheral oxygen release to tissues, has been found to be a valid concept during ALI/ARDS. In the most severe ARDS cases, extracorporeal membrane oxygenation has been used with success.

Death and organ transplantation

The diagnosis, confirmation and certification of death are core medical skills. This might sound obvious and straightforward, but in times when resuscitation techniques are more successful than ever and the medical benefits of organ transplantation are becoming widely recognized, special challenges for the clinician can arise.

Owing to mechanical ventilation, respiratory death can be prevented and has transformed the course of many terminal neurological disorders. This has led to the understanding that, along with the definition of cardiorespiratory death, a new definition has become necessary. Whereas in cardiorespiratory arrest the cessation of circulation will obviously lead to irreversible loss of brain function, the irreversible loss of brain function *with maintenance of circulation* has proved a novel concept. To date, medical professionals as well as lay people throughout the world have adopted the concept that a person is dead when his/her brain is dead.

This section will outline the principles of the diagnosis of death, with a special emphasis on brain stem testing and the principles of organ donation.

Death after cardiorespiratory arrest

There are currently no standardized criteria in the confirmation of death following irreversible cessation of cardiorespiratory function. As a result, practice still varies worldwide from confirming death as soon as the heart stops or when attempts at cardiopulmonary resuscitation are abandoned to waiting for 10 minutes or longer after the onset of asystole and apnoea. In the primary care setting this practice continues to be appropriate, whereas in hospitals the increasing practice of non-heart-beating organ donation warrants a standardized approach to confirming death. Guidance from the Academy of Medical Royal Colleges in the UK proposes that death can be diagnosed and confirmed after a minimum of 5 minutes of continuous observed asystole. Asystole can be identified by the absence of a central pulse on palpation and absence of heart sounds on auscultation. In the hospital and especially ICU setting, at least one of the following additional monitoring modalities should be used to prove absence of cardiac function: ECG display (also from a defibrillator), intra-arterial pressure monitoring and echocardiography. The time of death is recorded as the time at which these criteria are fulfilled.

Diagnosing brain death

Different protocols have been adopted among countries around the world for brain death testing and the requirements for number of observers, the specialty of the assessing physician, the duration and the use of confirmatory tests may differ. This section cannot account for all these differences. Requirements for special considerations are necessary for diagnosing brain death in children, an area that is reserved for the specialized physician. We will, however, focus on the principles of diagnosing brain death in adults. As a principle, because of conflict of interest, no member of the transplant team can be involved in brainstem testing.

As brain death occurs, reflexes will disappear in a rostral to caudal direction and the medulla oblongata is the last part of the brain to cease function. Several hours may be required for the destruction of the brainstem to be complete. The clinical neurological examination remains the standard for determination of brain death and has been adopted by most countries. This process must be carried out by experienced physicians and with precision. A complete clinical neurological examination includes documentation of coma, the absence of brainstem reflexes and apnoea.

Irreversible coma

First, the cause for persistent coma needs to be established and any reversibility must be excluded. Potential reversible factors are intoxication by drugs and medication, neuromuscular blocking agents, hypothermia and electrolyte imbalance (Table 6.10).

Acute endocrine imbalance associated with brainstem death is not the cause of reversible coma; chronic profound endocrine failures can usually be excluded from the history. Prolonged action and accumulation of hypnotics, tranquilizers and sedatives must be taken into account and sufficient time should be allowed for their clearance. If significant diagnostic uncertainty remains, brainstem testing cannot be undertaken.

Irreversible cessation of brainstem function

A thorough neurological examination must follow the assessment of irreversible coma to assess brainstem function (Table 6.11). To diagnose brainstem death, pupils must be assessed for absent reaction to bright light; a corneal reflex must be absent. No motor response to painful stimuli must be found. As reflexes at a spinal level might still be present, a *peripheral* motor response to *peripheral* stimulation can be observed, which is not a higher

Table 6.10 Potential reversible causes of coma that need to be addressed before testing for brainstem death

Factor	Lower limit	Upper limit	Comment
Temperature	34°C		Impaired consciousness <34°C, brainstem areflexia <28°C
Biochemistry			
Sodium	115 mmol/L	160 mmol/L	Derangements that are clearly the result of brain death (hypernatraemia associated with diabetes insipidus) may not require correction ahead of testing
Potassium	2 mmol/L	–	
Magnesium	0.5 mmol/L	3 mmol/L	
Phosphate	0.5 mmol/L	3 mmol/L	
Glucose	3 mmol/L	20 mmol/L	Blood glucose should be checked immediately before testing
Other parameters:			
pH	7.35	7.45	
$PaCO_2$	2.0 kPa	6.0 kPa	
PaO_2	10 kPa		
Mean arterial pressure	60 mmHg		

Table 6.11 Clinical criteria for brain death

Absence of all of the following:
• pupillary response to light
• corneal reflex
• motor response
• caloric response (oculovestibular reflex)
• gag reflex
• coughing in response to tracheal suctioning
• respiratory drive with hypercapnia

function of the brain. Therefore, application of painful stimuli in the distribution of the cranial nerves is preferred (supraorbital nerve, temporomandibular joint) and a lack of motor response must be observed. To test oculo-vestibular response, a clear view of the tympanic membrane should be established before instilling 50 mL of ice-cold water in each ear with the head tilted 30° to the horizontal plane. No eye movement must be observed following water instillation. Cough reflex to tracheal stimulation with a suction catheter must be absent as well as a gag reflex following stimulation of the posterior pharynx with a spatula. All of these tests must show absence of brainstem function before the apnoea test can be performed.

The apnoea test is the process of testing the respiratory response of the brainstem to hypercarbia and should be performed last. The patient should be in cardiovascular stability and ventilated to normocapnia before beginning the apnoea test. The apnoea test is not valid when direct high cervical injury was sustained (see below). The patient is disconnected from the mechanical ventilator, and observed for spontaneous respiratory efforts for 5 minutes. During the test, oxygen is inflated in the tracheal tube via a suction catheter to minimize hypoxia. Hypercarbia must be documented with blood gas analysis. After documenting the absence of respiratory response to hypercarbia, the patient should be reconnected to the ventilator for normoventilation.

Brainstem death can only be declared if all tests have failed to show any brainstem function. Most countries will acknowledge a patient's death after the first diagnosis of brainstem death. A second, confirmatory, brainstem test is often required before further management of the patient can be initiated.

Additional investigations and special considerations

The accuracy of clinical examination for the diagnosis of death by cessation of brainstem function has proven to be highly accurate and reliable. Therefore, not including results of neurophysiological and imaging investigations as part of these criteria has become widely accepted. However, death cannot be diagnosed on the basis of brainstem testing alone in circumstances when comprehensive neurological examination is not possible (such as extensive maxillofacial trauma, high cervical cord trauma or residual sedation). In such circumstances, confirmatory tests may reduce elements of uncertainty and shorten the period of observation prior to clinical brainstem testing.

Confirmatory tests can be categorized as assessing perfusion and neurophysiology:

- assessment of blood flow in the larger cerebral arteries (four-vessel angiography, transcranial Doppler, magnetic resonance angiogram and spiral CT angiogram)
- assessment of brain tissue perfusion (xenon CT, positron emission tomography, cerebral scintigraphy)
- assessment of neurophysiology (electroencephalogram, evoked potentials).

Not all tests are readily available in every hospital, have the same amount of diagnostic accuracy and reliability, nor are easy to interpret. In addition, haemodynamic instability occurs frequently in the dying patient, and transfer out of the ICU may not be without risk. When appropriate, such cases should be referred for a specialist neurological or neurosurgical opinion.

Differential diagnosis of brainstem death

Besides hypothermia and drug intoxication, differential diagnosis of brainstem death includes locked-in syndrome and Guillain–Barré syndrome.

Locked-in syndrome typically occurs following a destruction of the base of the pons. This is often caused by an acute embolus to the basilar artery. Patients with locked-in syndrome are conscious but cannot move their limbs, grimace or swallow, but voluntary blinking and vertical eye movements persist. Guillain–Barré syndrome is often reversible and can involve peripheral and cranial nerves. It represents an acute inflammatory demyelinating polyneuropathy, a disorder affecting the peripheral nervous system; patients have ascending paralysis, with weakness beginning in the feet and hands and migrating towards the core. Guillain–Barré syndrome is usually triggered by an acute infection and the progression occurs over a period of days. Meticulous history-taking should prevent the misdiagnosis of brain death.

Further management

After the clinical criteria of brainstem death have been met, the physician needs to inform the next of kin, who can be approached about organ donation. This can be done with the involvement of an organ procurement team. If organ donation is declined, withdrawing of ventilatory assistance may be the most appropriate course. However, if organ donation is considered a realistic option by the next of kin and the medical team, suitability of the patient should be confirmed by additional tests. This is usually a time-consuming process during which the patient must be treated appropriately. Normal body homeostasis should be maintained with fluid replacement and vasopressor/inotropic therapy, and ventilation must be continued. Electrolyte imbalances such as hypernatraemia, hypokalaemia and changes in glucose balance are frequent, especially if diabetes insipidus occurs. These should be corrected with the aim of normal ranges, as normalization of electrolytes will positively affect outcome of the transplanted organs. Substitution of hydrocortisone, thyroxine and antidiuretic hormone will help to balance the failure of the endogenous neuroendocrine response.

GUIDE TO FURTHER READING

Academy of Medical Royal Colleges. *A Code of Practice for the Diagnosis and Confirmation of Death*. London, UK: AOMRC, 2008. Available from: www.aomrc.org.uk/publications/reports-a-guidance.html.

Annane D, Sébille V, Charpentier C, *et al.* Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. *JAMA* 2002;**288**:862–871.

Bauby J-D. *The Diving Bell and the Butterfly: a Memoir of Life in Death*. New York, NY: Vintage Books, 1998.

Brealey D, Brand M, Hargreaves I, *et al.* Association between mitochondrial dysfunction and severity and outcome of septic shock. *Lancet* 2002;**360**:219–223.

Carré JE, Orban JC, Re L, *et al.* Survival in critical illness is associated with early activation of mitochondrial biogenesis. *Am J Respir Crit Care Med.* 2010;**182**:745–751.

Grocott MP, Martin DS, Levett DZ, *et al.*; Caudwell Xtreme Everest Research Group. Arterial blood gases and oxygen content in climbers on Mount Everest. *N Engl J Med* 2009;**360**:140–149.

Harvey S, Harrison DA, Singer M, *et al.* Assessment of the clinical effectiveness of pulmonary artery catheters in management of patients in intensive care (PAC-Man): a randomised controlled trial. *Lancet* 2005;**366**:472–477.

National Confidential Enquiry into Patient Outcome and Death. An acute problem? A report of the National Confidential Enquiry into Patient Outcome and Death. London, UK: NCEPOD, 2005. Available from: http://www.ncepod.org.uk/2005report/introduction.html.

Peek GJ, Mugford M, Tiruvoipati R, *et al.* for the CESAR Trial Collaboration. Efficacy and economic assessment of conventional ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory failure (CESAR): a multicentre randomised controlled trial. *Lancet* 2009;**374**:1351–1363.

Pronovost PJ, Angus DJ, Dorman T, *et al.* Physician staffing patterns and clinical outcomes in critically ill patients. *JAMA* 2002;**288**:2151–2162.

Ranieri VM, Thompson BT, Barie PS, *et al.*; PROWESS-SHOCK Study Group. Drotrecogin alfa (activated) in adults with septic shock. *N Engl J Med* 2012;**366**:2055–2064.

Rivers E, Nguyen B, Havstad S, *et al.*; Early Goal-Directed Therapy Collaborative Group. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001;**345**:1368–1377.

Shoemaker WC, Appel PL, Kram HB, *et al.* Prospective trial of supranormal values of survivors as therapeutic goals in high-risk surgical patients. *Chest* 1988;**94**:1176–1186.

Soni N. ARDS, acronyms and the Pinocchio effect. *Anaesthesia* 2010;**65**:976–979.

Sprung CL, Annane D, Keh D, *et al.*; CORTICUS Study Group. Hydrocortisone therapy for patients with septic shock. *N Engl J Med* 2008;**358**:111–124.

Suter PM, Fairley B, Isenberg MD. Optimum end-expiratory airway pressure in patients with acute pulmonary failure. *N Engl J Med* 1975;**292**:284–289.

van den Berghe G, Wouters P, Weekers F, *et al.* Intensive insulin therapy in critically ill patients. *N Engl J Med* 2001;**345**:1359–1367.

West JB. *Respiratory Physiology: The Essentials.* Baltimore, MD: Lippincott Williams & Wilkins, 2008.

Wijdicks E. The diagnosis of brain death. *N Engl J Med* 2001;**344**: 1215–1221.

CHAPTER 7

Anaemia, hypovolaemia and oedema

SIR ALFRED CUSCHIERI

Introduction

This chapter covers the pathophysiology and management of patients who have a reduced haemoglobin (Hb) load (anaemia), or who sustain a sudden reduction in the circulatory blood volume (hypovolaemia) or who accumulate abnormal amounts of fluid in the interstitial space or serous cavities (oedema). Transfusion of blood and blood products is also covered in this chapter. The spectrum of disorders covered is wide and of variable surgical urgency, from the patient with iron-deficiency anaemia requiring elective surgery for a right-sided colon cancer to the emergency management of a patient with profound hypotension from major ongoing blood loss, and patients with generalized oedema due to cardiac, hepatic or renal disease.

Anaemic patients: types, pathophysiology and management

Definition and types of anaemias

Anaemia is defined as a reduced level of circulating Hb. Minor degrees are asymptomatic as the oxygen-carrying capacity of the blood, although reduced, is sufficient to meet the needs for daily activities. Below a certain level of Hb (8.0 g/dL), the oxygen-carrying capacity is sufficiently compromised as to be unable to meet the oxygen needs except at rest and the patient becomes symptomatic (symptomatic anaemia). The primary symptoms are those of oxygen lack (asthenia, lethargy, tiredness, diminished exercise, tolerance, etc.); others follow from the compensatory mechanism that consists essentially of a fast hyperkinetic circulation giving rise to palpitations, buzzing in the ears and angina. Severe anaemia may indeed cause heart failure. The types of anaemia are summarized in Box 7.1.

BOX 7.1 Types of anaemia

- Iron deficiency (hypochromic microcytic)
- Sideroblastic (hypochromic microcytic + ringed sideroblasts in the marrow)
- Anaemia of chronic disease (normocytic normochromic)
- Macrocytic normoblastic
  - Liver disease
  - Alcoholism
  - Reticulocytosis secondary to haemolysis
- Macrocytic megaloblastic
- Congenital haemolytic (various morphological types, splenomegaly)
  - Red cell membrane defects: hereditary spherocytosis, hereditary elliptocytosis
  - Red cell enzyme defects: glucose-6-phosphate deficiency, pyruvate kinase deficiency
  - Haemoglobinopathies (abnormal haemglobins): sickle cell disease, thalassaemia
- Acquired haemolytic
  - Idiopathic warm autoimmune haemolysis [active at 37°C, immunoglobulin (Ig) G antibodies]
  - Idiopathic cold autoimmune haemolysis (active at lower temperatures, IgM antibodies)
  - Drug-induced immune haemolysis
  - Haemolytic disease of the newborn (rhesus D incompatibility)
- Aplastic/hypoplastic (bone marrow stem cell failure/depression)

In surgical practice, by far the most common is iron-deficiency (hypochromic microcytic) anaemia due to chronic occult blood loss (low serum ferritin and iron, high serum transferrin with low percentage saturation) exemplified by carcinoma of the right colon or stomach. In right colon cancer, hypochromic microcytic anaemia with positive faecal occult blood is the only early manifestation of the disease.

Megaloblastic anaemia arises from deficiency of vitamin B_{12}, folate or both. In a surgical context, megaloblastic anaemia is important from the preventive aspect, i.e. the need for vitamin B_{12} replacement therapy after gastrectomy and distal ileal resection/disease; it occurs (due to folate deficiency) in patients on chemotherapy for cancer and in those on chronic dialysis prior to renal transplantation. Pernicious anaemia (megaloblastic anaemia + atrophic gastritis with achlorhydria + parietal cell antibodies and impaired absorption of vitamin B_{12} corrected by oral intrinsic factor) assumes surgical importance because of the high risk for gastric cancer in these patients.

Anaemia associated with chronic disorders (normocytic normochromic) is commonly encountered in surgical practice in patients with malignant disease. It often does not respond to haematinics and then requires blood transfusion. Iron-deficiency anaemia and the anaemia of chronic disease may coexist. Patients with marrow failure (leukaemia, high-dose chemotherapy and malignant infiltration) require haematological support with blood transfusion, other component transfusion (platelets, granulocytes, clotting factors) and marrow rescue agents (recombinant colony-stimulating factors).

Anaemia and blood volume

The blood transfusion requirements of the anaemic patient are very different from those of the patient with acute blood loss. The crucial difference between anaemia and hypovolaemia (from acute haemorrhage) concerns the circulating blood volume (CBV), which is maintained or increased in anaemic individuals irrespective of the morphological type and aetiology of the anaemia (even when due to chronic blood loss). The deficit in anaemia thus relates to the Hb load and hence to the oxygen-carrying capacity of the blood. This reduced Hb level in symptomatic anaemia, even when fully saturated (oxygen saturation = 100%), is insufficient to provide normal tissue oxygenation. The compensatory mechanism consists of a hyperkinetic circulation with tachycardia and fast circulation time. In itself, this persistent tachycardia may contribute to anaemic cardiac failure due to deficient oxygenation of the myocardium (reduced diastolic interval).

During transfusion of anaemic patients, therefore, the prime consideration from the therapeutic safety standpoint is the prevention of circulatory overload and congestive cardiac failure. The objective of the transfusion is thus to raise the Hb load in a volume and during a time scale (slow transfusion over a 4 hour period) that can be tolerated by the patient. A transfusion of red cells of 4 mL/kg will raise the circulating Hb concentration by approximately 1.0 g/dL. Thus, each unit of red cells (300 mL) should raise the circulating Hb in a 70 kg adult by this amount. As the reason for transfusion of anaemic patients is to increase the oxygen-carrying capacity of blood, the red cell concentrates should be less than 14 days old as these have near normal levels of 2,3-diphosphoglycerate (2,3-DPG).

Nature, causes and consequences of hypovolaemia

Hypovolaemia is defined as a reduced CBV (= 70 mL/kg in adults; 80 mL/kg in infants) and can be either true or apparent. True hypovolaemia results from contraction of the CBV as a result of losses (blood from haemorrhage, plasma in the burnt patient and dehydration from deficits of water and saline). Apparent hypovolaemia ensues from an increased vascular capacity, usually due to loss of the peripheral resistance in the muscular arterioles (sepsis, adrenal insufficiency, anaphylaxis, neurogenic factors), although in some of these conditions, e.g. sepsis, the situation is more complex as the increased capillary permeability induces intravascular fluid losses into the interstitial space.

Response to hypovolaemia

Ultimately, the adverse effects of hypovolaemia are due to an inadequate (not necessarily reduced) cardiac output (CO) and hence inadequate cellular perfusion. To appreciate the pathophysiological consequences of hypovolaemia, one has to appreciate the normal control of CO. Overall CO is determined by the stroke volume (SV) and heart rate (HR) as follows: CO = SV × HR. In a young, fit adult at rest, each ventricle fills during diastole to reach an end-diastolic volume of 120 mL of blood. With each ventricular contraction (systole), 70 mL is ejected from each ventricle. The actual amount is known as the ejection fraction, and is normally 60% of the volume of blood present in each ventricle at the end of diastole. In accordance with Starling's law (force of cardiac contraction increases with the diastolic stretch of the cardiac muscle fibres), the more the heart fills during diastole, the greater the amount of blood expelled with each beat (SV), but the ejection fraction remains fairly constant. CO is then determined by (1) preload, (2) cardiac contractility, (3) afterload and (4) HR.

Preload

This refers to venous filling, venous return and is determined by the CBV and venous tone. The peripheral veins, especially the venules of the liver and spleen (capacitance vessels), act as a large reservoir and contain 45% of the intravascular volume, whereas the large central veins hold about 18%. Hence, preload is reduced significantly in hypovolaemia. To some extent, the compensatory venous vasoconstriction due to the release of norepinephrine (noradrenaline) from the sympathetic nerve endings in the vein walls may help to maintain an adequate preload up to a certain deficit in the CBV

Two other mechanisms facilitate the venous return: respiratory excursions and skeletal muscle activity. During inspiration, the diaphragm descends, increasing the intra-abdominal pressure that is transmitted to the intra-abdominal veins. At the same time, the intrathoracic pressure falls, lowering the right atrial pressure, the net effect being to accentuate venous return.

This mechanism remains operative in hypovolaemia, unless complicated by chest injuries, which preclude normal ventilation (flail chest, haemopneumothorax and ruptured diaphragm). The contraction of the skeletal muscles within their fascial compartments is the main contributor of venous return in the lower limbs (the muscle pump). The contracting muscle bulk restrained by the overlying fascial sheath directly compresses the deep veins and encourages flow from the superficial to the deep venous systems. The efficiency of the muscle pump is dependent on the integrity of the venous valves. The muscle pump is lost in the prostrated hypovolaemic patient, but, because of the supine or head-down position in which these patients are nursed, the loss of the muscle pump is not an important factor in the reduced preload.

Cardiac contractility

Cardiac contractility is diminished in hypovolaemia as a consequence of the reduced end-diastolic filling/volume of the ventricles (and thus stretch of the cardiac muscle fibres).

Afterload

Afterload equals the peripheral arterial resistance which acts as an impediment to cardiac ejection and thus a reduced afterload (e.g. by vasodilator therapy) tends to favour an increase in the CO, but only in the presence of an adequate CBV. In the hypovolaemic patient, peripheral resistance is initially increased (sympathetic and adrenal response) so that the percentage fall in the blood pressure is always an underestimate of the drop in the CO. Thus, for example, in previously healthy young adults, the acute loss of 20% of the blood volume reduces, on average, the arterial pressure by 15% and the CO by 41%.

Heart rate

Within limits, tachycardia increases the CO but extreme HRs reduce output by diminishing ventricular filling and coronary artery nutrient blood flow to the myocardium.

The response of an individual to blood loss varies considerably. Indeed, some young, healthy individuals, by virtue of the intense vasoconstriction they can mount in response to volume deficits, can lose as much as 25% of the CVB without any significant change in the arterial blood pressure. The response is influenced by age, duration and severity, pre-existing myocardial disease, anaemia and associated trauma. In the accident victim, the additional trauma alters the neuroendocrine and metabolic response. Thus, trauma victims with mild to moderate hypovolaemia may have a normal or even an elevated CO, and they may be normotensive or even hypertensive after the accident. Even the HR response is variable in these patients, with either a tachycardia or bradycardia at the time of resuscitation.

Consequences of hypovolaemia

Diminished tissue perfusion, tissue hypoxia, anaerobic glycolysis and lactic acidosis

The immediate consequence is hypovolaemic shock, defined as a state of inadequate cellular perfusion. The delivery of oxygen (DO_2) to the tissues is reflected by the product of CO × arterial oxygen content (CaO_2) adjusted to the patient's build (indexed):

$$DO_2 \text{ (mL/min/m}^2) = \text{CO} \times CaO_2 \times 10$$

The arterial oxygen content depends on the Hb level and the percentage oxygen saturation. Under normal physiological conditions, DO_2 ranges from 500 to 720 mL/min/m^2 and the oxygen consumption (VO_2) between 100 and 160 mL/min/m^2, representing an oxygen extraction ratio of 22–30%. This level of tissue oxygenation ensures aerobic metabolism of glucose with the production of ATP, CO_2 and H_2O via pyruvate so that the blood lactate is very low (>2.0 mmol/L) unless there is sustained muscular exertion when a certain amount of anaerobic metabolism is incurred with the production of lactate. This tendency to acidosis shifts the oxyhaemoglobin to the right, favouring dissociation, i.e. increased release of oxygen. Thus, any fall in the DO_2 is compensated for by increased extraction of oxygen. Tissue oxygenation is disturbed in all forms of shock.

DO_2 is reduced in hypovolaemic shock and is accompanied by an increased extraction, so that VO_2 is not markedly depressed unless the hypovolaemia is extreme. The severity of the lactic acidosis and the extent of reduction of the mixed venous Hb saturation in the pulmonary artery, SVO_2 (obtained by pulse oximetry of a blood sample or from the distal lumen of a Swan–Ganz catheter), depend on the severity of the hypovolaemia and on adequate early restoration of the blood volume.

Acute renal failure

Every organ incurs hypoxic damage from sustained hypovolaemia. The intense peripheral and splanchnic vasoconstriction that occurs in response to blood volume losses to redistribute the available blood to the brain puts the kidneys at particular risk of ischaemic damage: acute renal failure (ARF). Initially (prerenal or potential ARF), the renal function is simply responding to the prerenal deficits and intense vascular shut down, with low perfusion pressure and reduced oxygen delivery. There is 'physiological' oliguria with the production of small volumes of concentrated urine and low fractional sodium excretion. Renal tubular function is relatively well maintained. Urine osmolality is still greater than serum osmolality (urine/plasma osmolality ratio >1.05). There are no structural changes and recognition of the problem at this stage with adequate correction of the prerenal deficits can prevent the development of established acute tubular necrosis (ATN). ARF (pathologically ATN) is a potentially reversible reduction in the capacity of the kidney to excrete nitrogenous wastes and maintain fluid and electrolyte homoeostasis. ARF is common, complicating 5% of all medical and surgical admissions in a large American study. A UK community-based study reported an overall annual incidence of ARF of 172 cases per million adult population.

Pathogenesis of acute tubular necrosis

Overall the renal blood flow amounts to 25% of CO but most of this supplies the cortex, which contains the glomeruli and convoluted tubules as these require preferential perfusion for filtration and reabsorption, the latter having high energy requirements. In contrast the outer medulla receives blood which is low in oxygen hydrostatic pressure after having traversed the glomerular capillary bed. These features, though essential for maintaining the osmotic gradients within the medulla, render it susceptible to variations in blood flow especially as the medulla contains the ascending limb of the loop of Henle and the S3 segment of the proximal tubule, which have high oxygen requirements. The resulting impaired tubular sodium reabsorption encountered is sustained in hypovolaemia as the reduced perfusion causes constriction of the afferent arterioles, which further reduces the glomerular filtration rate. This tubular–glomerular feedback (designed to protect the downstream nephron) may cause renal injury if the low-flow state is prolonged. The reduced blood flow in the peritubular capillaries produces ischaemic damage in vascular endothelial cells, resulting in cell swelling and leucocyte activation, which in turn produce cytokines and reactive oxygen species that damage endothelial and tubular epithelial cells. Tubular cells swell, lose their brush border and develop cytoskeletal abnormalities with abnormal localization of cell membrane components (Na$^+$/K$^+$-ATPase) and loss of cell–cell and cell–basement membrane attachment. These swollen cells detach and obstruct the tubular lumen. The exfoliation of tubular cells is thought to be due to redistribution of the integrin adhesion molecule receptors from the basolateral to the apical membranes. The exfoliated cells adhere downstream to other tubular cells to form casts. The ATN tubular damage is thus found predominantly in the outer medulla where oxygen delivery is always borderline. Changes in the cortical tubules are less marked, partly because cortical blood flow, and hence transport workload, fall rapidly after the ischaemic insult. The renal impairment results from (1) a fall in the whole kidney glomerular filtration rate, (2) obstruction of tubules by cellular debris/casts and (3) back-leakage of filtrate across the damaged tubules into the interstitium. Histologically in ATN, the tubules are surrounded by flattened, denuded epithelium, and the lumen filled by cell debris, with congested peritubular capillaries, extensive inflammatory cell infiltrate and cell necrosis

A previously healthy patient is able to regain normal structure and function after ATN if renal perfusion and oxygen supply are normalized. The regeneration is from viable cells adherent to the tubular basement membrane, which divides and spreads to cover denuded areas before differentiation to normal tubular epithelium and function. The return of glomerular filtration aids clearance of tubular debris and relief of obstruction. Often there is a period when glomerular filtration has returned to normal before tubular function is restored. This accounts for the polyuric phase of ATN.

Risk, Injury, Failure, Loss and End-stage Kidney (RIFLE) classification

To establish a uniform definition for acute kidney injury, the Acute Dialysis Quality Initiative formulated the Risk, Injury, Failure, Loss, and End-stage Kidney (RIFLE) classification (Table 7.1). RIFLE defines three grades of increasing severity of acute kidney injury – *risk* (class R), *injury* (class I) and *failure* (class F) – and two outcomes: (1) loss (persistent ARF = complete loss of kidney function >4 weeks) and (2) end-stage kidney disease. The RIFLE classification provides three grades of severity for acute kidney injury based on changes in either serum creatinine or urine output from the baseline condition into one of the three RIFLE severity classes: acute kidney risk (R), acute renal injury (I) or ARF (F). Within the general intensive care unit population, acute kidney 'risk, injury, failure', as defined by the RIFLE classification, has been shown to be associated with increased hospital mortality and patients in the RIFLE class R are at high risk of progression to class I or class F. RIFLE class I or class F patients have an increased risk of in-hospital mortality compared with those who do not progress beyond class R irrespective of severity of illness, case mix, sex and age.

Table 7.1 Risk, Injury, Failure, Loss, and End-stage Kidney (RIFLE) classification

Category	GFR criteria	UO criteria	
Risk	Increased creatinine × 1.5 or GFR decrease >25%	UO <0.5 mL/kg/h × 6 hours	High sensitivity
Injury	Increased creatinine × 2 or GFR decrease >50%	UO <0.5 mL/kg/h × 12 hours	
Failure	Increase creatinine × 3 or GFR decrease >75%	UO <0.3 mL/kg/h × 24 hours or anuria × 12 hours	High Specificity
Loss	Persistent ARF = complete loss of kidney function >4 weeks		
ESKD	End-stage kidney disease (>3 months)		

ARF, acute renal failure; ESKD, end-stage kidney disease; GFR, glomerular filtration rate; UO, urine output.

Recovery of acute reversible intrinsic renal failure

Clinically, recovery of sufficient renal function for independent survival after oliguric acute reversible intrinsic renal failure takes 2–4 weeks. Old age, sepsis, jaundice and cardiorespiratory failure are adverse factors that usually prevent recovery of renal function.

Management of acute renal failure

ARF develops in up to 5% of patients at hospital admission and affects up to 20% of patients in intensive care units. The aetiology in the individual patient can be identified in most instances by an appropriate history, physical examination and selected laboratory investigations. Management is aimed in the first instance at correction of life-threatening consequences and providing support of the patient until recovery of renal function. The life-threatening consequences of ATN are (1) hyperkalaemia, (2) pulmonary oedema and (3) severe acidosis and these require immediate attention. Additionally, any causative factors, including nephrotoxic drugs, that may have contributed to ATN must be withdrawn and corrective measures put in place and any urinary tract obstruction must be relieved. Renal replacement therapy (RRT) is used when appropriate.

Hyperkalaemia

Severe hyperkalaemia (plasma potassium >6.5 mmol/L) is a medical emergency because of the risk of cardiac arrhythmias and arrest. Severe hyperkalaemia is accompanied by a typical pattern of ECG changes: peaked/tented T waves, flattening of the P wave and prolongation of the QRS complex, sine waves with higher K levels (8–9 mmol/L) ventricular fibrillation or asystole. Urgent treatment of hyperkalaemia must be initiated when the serum potassium exceeds 6.5 mmol/L, or in the presence of any ECG changes.

The treatment of hyperkalaemia entails:

- Stabilization of cardiac myocytes by calcium as this antagonizes the effects of hyperkalaemia on the myocardium. Calcium should be administered as soon as P wave or QRS changes are noted on the ECG as a slow intravenous bolus of 10–20 mL of 10% calcium gluconate over 2–5 minutes. In patients on digoxin, calcium is administered as a slow infusion (10 mL 10% calcium gluconate in 100 mL 5% dextrose over 30 minutes) as calcium potentiates cardiac toxicity of digoxin.
- Measures to reduce plasma potassium concentration: insulin with glucose to increase cellular potassium uptake (by activation of Na^+/H^+ channels). The addition of glucose to the insulin bolus is necessary to prevent hypoglycaemia. In practice, 10 units of fast-acting soluble insulin are added to 50 mL of 50% dextrose and infused over 10–20 minutes. Insulin alone can be given to hyperglycaemic patients (blood glucose >14 mmol/L). During this treatment, the blood glucose should be monitored every 6 hours to ensure against hypoglycaemia.
- Removal of potassium from the body: ion exchange resins bind potassium in the gastrointestinal (GI) tract, in exchange for calcium or sodium, with resulting increased potassium excretion in the stool. Calcium resonium and Resonium A are commonly used, in an oral dose of 15 g t.d.s. together with an osmotic laxative. Ion exchange resins are most effective when potassium levels are moderately high (6.5 mmol/L).
- Haemodialysis is indicated in refractory severe hyperkalaemia and is more effective than haemofiltration or peritoneal dialysis. Maximum removal occurs in the first hour of dialysis. Serum potassium may decrease by 1.2–1.5 mEq/L/h if potassium free dialysate is used, which can precipitate hypokalaemia and arrhythmias. Generally a dialysate potassium concentration of 1.0 or 1.5 mmol/L is used.

Pulmonary oedema

This results from fluid overload and is usually accompanied by ventilatory failure, which, if significant, requires supplementary oxygen and non-invasive ventilation, or intubation and ventilation if severe. Pharmacological treatment consists of intravenous infusion of glyceryl trinitrate (50 mg in 50 mL 0.9% saline, at a rate of 2–20 mL/h) aimed at keeping the systolic blood pressure >95 mmHg) and administration of loop diuretics in large doses (furosemide 250 mg in 50 mL 0.9% saline) to achieve diuresis. If these interventions prove ineffective, RRT, haemodialysis or haemofiltration is indicated.

Acidosis

Severe metabolic acidosis (blood pH <7.2) is frequently present and used to be treated with the administration of sodium bicarbonate, although there is no evidence that this is beneficial and may be counterproductive as it increases the sodium load. Much lower concentrations – isotonic (1.26%) solutions – of sodium bicarbonate may have a role as fluid replacement therapy in stable patients with a moderate to severe acidosis who require fluid replacement and in whom the need for dialysis is not imminent. Otherwise haemodialysis or haemofiltration is required for the treatment of severe acidosis in oligoanuric patients.

Fluid balance

The optimal fluid is still debated. Normal saline and 4% albumin have clinical equivalence, as shown by a study of patients admitted to intensive care units. Once the patient is considered normovolaemic, care should be taken to avoid fluid overload and a maintenance regimen started, with careful fluid balance charts to take into account renal and insensible fluid losses, aimed at achieving a positive balance not exceeding 500 mL/day.

Relief of obstruction

Any obstruction of the urinary tract requires prompt relief. For bladder outflow obstruction, a urethral or suprapubic catheter is used. Either is considered necessary in all patients with ARF for accurate measurement of urine output. Relief of upper tract obstruction may require antegrade (percutaneous nephrostomy) or retrograde (cystoscopy and retrograde ureteral catheterization) approaches for the relief of the obstruction. A significant diuresis can complicate relief of complete urinary tract obstruction and has to be taken into consideration in the fluid balance.

General measures

Patients with ARF may develop a bleeding tendency from platelet dysfunction and coagulopathies. Correction of coagulopathy is with blood products and vitamin K. RRT may improve platelet function but desmopressin (DDAVP) and cryoprecipitate may be required. Carbohydrate and protein requirements should be tailored individually and ideally delivered via the enteral route, although parenteral administration may be necessary in some patients. Patients with ARF are also more susceptible to infections, which often result in clinical deterioration; infection control measures should be instituted. Finally, prophylaxis against deep venous thrombosis is indicated in patients with ARF.

Renal replacement therapy

The indications for RRT are:

- severe hyperkalaemia, unresponsive to medical therapy
- fluid overload with pulmonary oedema
- uraemia (blood urea >30–50 mmol/L)
- complications of severe uraemia: encephalopathy, pericarditis, neuropathy/myopathy
- severe acidosis (pH <7.1)
- drug overdose with a dialysable toxin.

Management of hypovolaemia

The management of the hypovolaemic patient entails:

- detection of the hypovolaemia, its cause and severity
- arrest of haemorrhage
- establishing good intravenous access with large-bore peripheral cannulas (two if hypovolaemia is severe)
- replacing the deficit
- repeated clinical observations and monitoring the patient.

Detection of hypovolaemia, its cause and severity

Although in the majority of patients the diagnosis of hypovolaemia is obvious, in some patients it is less clear-cut because of previous partial resuscitation and variable degrees of vasoconstriction. Diagnosis and assessment are thus based on clinical assessment, including careful history, clinical examination, estimation of observed losses, haemodynamic measurements (HR, pulse volume, arterial and venous pressures and haematocrit) and indices of tissue perfusion (nail bed capillary refill, peripheral skin temperature, urine output and level of consciousness). It must be stressed that the haemodynamic measurements are complementary to the clinical assessment, as a normotensive patient with a normal HR may still be hypovolaemic.

The severity of hypovolaemia correlates with the volume lost and is reflected in the clinical signs. In general, contraction of the CBV by more than 25% (1.5 litres) is regarded as severe hypovolaemia. The severity of the hypovolaemia is much more difficult to assess clinically when the hypovolaemia is secondary to deficits of water and salt. In these instances, careful reappraisal of the fluid balance charts gives the best estimate of the losses incurred.

Intravenous access and replacement of deficits

The principle of volume replacement, 'replace like with like', needs to be qualified in respect of the cause of the hypovolaemia. It applies fully to hypovolaemia associated with fluid and electrolyte losses. Burn patients may require blood in addition to crystalloids and protein solutions. A previously healthy adult is able to tolerate a blood loss of 25% of the CBV (1.5 litres) with volume replacement by crystalloids or plasma expanders only. Beyond this loss, blood transfusion becomes necessary as the oxygen-carrying capacity of the blood is then compromised by the significant reduction in the Hb content of the blood.

Although good intravenous access for volume replacement is an obvious necessity, it is often compromised by inability to insert an appropriately sized cannula (or two in severe hypovolaemia) because of the collapsed veins. Valuable time is often lost in persisting with efforts to cannulate percutaneously a peripheral upper limb vein. In these cases, a cut-down is made over a suitable vein. In recent years there has been a shift away from the use of crystalloid solutions in the resuscitation of patients with hypovolaemia due to blood loss because of the established risks of tissue and especially pulmonary oedema and the development of compartment syndromes. Major blood losses should be replaced with red cell concentrates, platelet transfusions and clotting factor concentrates according to established protocols. These protocols for massive blood transfusions are now in established practice in all major trauma centres. All such protocols for massive transfusion now focus on the prevention of coagulopathy and thrombocytopenia. A 1:1:1 ratio [equal parts packed red blood cells (PRBCs), fresh frozen plasma (FFP) and platelets] for blood component therapy is now recommended as a more physiological regimen aimed at prevention of complications. This approach is called haemostatic resuscitation because it emphasizes the early correction of coagulopathy, which is conducive to improved survival.

Hypovolaemia due to acute blood loss

The management entails:

- Replacing the volume at the same time as controlling the bleeding. Blood samples are taken for blood group/antibody screening, urea and electrolytes and haematology, i.e. Hb, packed cell volume (PCV) and platelet count. Rapid intravenous infusion is commenced with crystalloid solutions, e.g. isotonic saline or colloids (protein solutions or non-plasma colloid volume expanders). Several clinical trials have confirmed that colloid solutions are not superior to crystalloids for resuscitation but much larger volumes of crystalloids are needed. Resuscitation with isotonic saline should be avoided in patients with severe liver disease in view of the risk of sodium overload. For the same reason, only salt-poor albumin solutions should be administered to these patients. Dextran and hetastarch [Hespan, hydroxyethyl starch (HES)] should be limited to 1.5 L/24 h.
- Blood transfusion if losses exceed 1.5 litres.
- Administering oxygen by facemask or nasal catheter.
- Requesting an early coagulation screen.
- Repeating clinical observation and monitoring of the patient.

Massive blood transfusion

Massive blood transfusion is defined as that which is equivalent to or exceeds the patient's own blood volume within 12 hours. It carries high morbidity and mortality rates, primarily because of the underlying condition that necessitates the massive blood transfusion and, to a lesser extent, because of the adverse acute changes that are caused by the rapid infusion of large amounts of cold stored blood. Patients requiring massive blood transfusion form a heterogeneous group: young, previously healthy trauma victims, patients with major bleeding disorders, obstetric complications, etc. The nature of the underlying condition and the age of the patient largely determine the survival. Pre-existing renal and liver disease and old age militate against recovery.

Aside from being cold, stored blood has an acid pH, contains citrate anticoagulant and has an elevated plasma K^+ and ammonia and a reduced 2,3-DPG. The metabolic consequences therefore include hypothermia, acidosis, hyperkalaemia and an increased affinity of the transfused Hb for oxygen (reduced 2,3-DPG), thereby contributing to the tissue hypoxia. The reduced oxygen-carrying capacity is not a real problem nowadays with the use of modern anticoagulant preservative solutions that ensure adequate concentration of 2,3-DPG in stored blood for up to 14 days. For this reason, blood that is less than 14 days old is recommended for massive blood transfusions. The reduced oxygen-carrying capacity is likely to be important only in anaemic patients and in those with pre-existing cardiac disease. Hypothermia may lead to cardiac arrhythmias including ventricular fibrillation and asystole. For this reason, patient warming (heating blankets, etc.) and blood warming are necessary if the transfusion rate exceeds 50 mL/min. Unfortunately, the heating coils increase the resistance of the oxygen-giving circuit, but their use is essential in these patients. Citrate intoxication is due to the chelation of ionized calcium, which may result in the prolongation of the QT interval, but this does not materially affect cardiac function and the ionized calcium level rapidly returns to normal after the transfusion as the excess citrate is metabolized and excreted. Thus, the use of supplemental calcium is not justified, particularly as it may itself give rise to arrhythmias. Hyperkalaemia is seldom a problem as the excess plasma K^+ enters the red blood cells (RBCs) with warming to body temperatures. It is, however, a consideration in patients with acidosis and renal failure when calcium is administered as the physiological antidote.

Clotting factors and other blood component deficiencies that arise as a result of massive blood transfusion are more important than the metabolic complications:

- clotting factor deficiency, especially factors V and VIII
- thrombocytopenia
- disseminated intravascular coagulation (DIC)
- reduced plasma colloid osmotic pressure.

Stored blood is deficient in factors V and VIII and beyond 48 hours contains practically no functioning platelets. The situation is further compounded by the dilution that occurs in these patients because of infusion of crystalloids and plasma expanders before or in between units of blood. The breakdown of some of the transfused cellular components (RBCs, platelets, leucocytes) releases thromboplastin-like material and thus can trigger DIC.

Acute respiratory distress syndrome (ARDS) may complicate massive blood transfusion and is often multifactorial. Its incidence is influenced by the underlying condition (sepsis and major trauma), but microemboli from white cell and platelet aggregates and reduced plasma oncotic pressure (dilution) can contribute to the development of the syndrome. ARDS can also develop as a result of transfusion-related acute lung injury (TRALI; see below).

The modern practice of 'safe' massive blood transfusion consists of:

- replacing and maintaining blood volume by RBC packs: use stored blood less than 14 days old
- haemostatic resuscitation by administering a 1:1:1 ratio (equal parts PRBCs, FFP and platelets); this has replaced maintaining haemostasis based on the coagulation screen (platelet count, thrombin time and prothrombin time)
- maintaining oxygen-carrying capacity: ensure PCV >20 or Hb >80 g/L (repeat haematological testing)
- keeping the patient warm: heating coils on the blood circuit, heated blanket, etc.
- correcting or avoiding metabolic complications
- maintaining a normal plasma protein concentration: avoid excessive crystalloid infusions and monitor the plasma albumin level.

DIC is presumed if the thrombin time is more than double the control value. In these cases, 10 units of cryoprecipitate are administered in addition to FFP to correct the deficiency of clotting factors (V, VIII).

Monitoring of the hypovolaemic patient

The primary objectives of monitoring are to ensure that the CBV has been replaced and that tissue perfusion has been restored to normal. In addition, in the unstable patient monitoring provides early signs of either renewed bleeding or cardiac decompensation and thus the need for inotropic support. The extent of monitoring needed depends on the severity of the hypovolaemia, associated comorbid cardiorespiratory disease or trauma and cardiovascular stability of the patient.

Monitoring of the otherwise stable patient who recovers quickly with volume replacement includes:

- arterial blood pressure, core temperature and pulse rate
- central venous pressure (CVP): if values are equivocal, the effect of a fluid challenge on the CVP (100–200 mL over 2–3 minutes) will determine whether the patient is still hypovolaemic (CVP remains unchanged or may even drop as the vasoconstriction subsides) or is likely to be overloaded (sharp elevation in CVP)
- hourly urine output: an output >30 mL urine/h indicates adequate renal perfusion
- pulse oximetry: preferably one that also displays the pulse plethysmogram in addition to percentage saturation of the Hb.

More intensive monitoring is only required in cardiovascularly unstable patients, including those who sustain major trauma. This includes (in addition to the above):

- insertion of a radial artery cannula for continuous monitoring of the arterial pressure and to obtain samples for blood gas analysis
- core–peripheral temperature gradient: a useful, non-invasive indicator of peripheral perfusion
- insertion of a pulmonary artery flotation catheter (Swan–Ganz catheter): this allows monitoring of the right atrial pressure, pulmonary artery pressure and pulmonary capillary wedge pressure, measurement of the CO (by thermodilution) and sampling of the mixed venous blood for oxygen saturation (SvO_2). A number of important derived variables can be obtained from these measurements in conjunction with the results of blood gas analysis. These are pulmonary vascular resistance, systemic vascular resistance, oxygen extraction ratio and systemic oxygen consumption.

Gastrointestinal haemorrhage

It is customary to consider GI bleeding as occurring from either the upper or the lower GI tract. This classification, although in established usage, is not entirely satisfactory as it omits a group of disorders, albeit less common, that present with bleeding from lesions in the midgut (from the duodenojejunal junction to the proximal transverse colon). In the upper/lower GI classification, right colonic bleeding lesions are grouped with lower GI bleeding. The causes of GI bleeding by topographical site in the gut are outlined in Table 7.2.

Table 7.2 Causes of gastrointestinal bleeding by topographical site

Foregut	Midgut	Hindgut
Oesophagitis, oesophageal varices	Ulcers: usually drug induced Small bowel tumours	Colorectal tumours/polyps Inflammatory bowel disease
Erosive gastritis	Vascular anomalies: right colon and small bowel	Diverticular disease
Peptic ulcer: DU, GU	Hereditary telangiectasia	Angiodysplasia
Tumours: adenocarcinoma, smooth muscle, lymphoma	Peutz–Jeger polyps	Ischaemic colitis
Vascular anomalies	Jejunal diverticula	Trauma
Hereditary telangiectasia	Meckel's diverticulum: infants and children	Endometriosis
Anastomotic suture line: postoperative	Aortoenteric fistula: patients with aortic grafts	Anastomotic suture line: postoperative
Mallory–Weiss syndrome	Crohn disease	Portal hypertension
Trauma: including iatrogenic	Anastomotic suture line: postoperative	Haemorrhoids
Haemobilia	Right colon cancer/polyps	Anal fissure
Chronic pancreatitis	Portal hypertension: includes bleeding from stomal varices	Bleeding disorder*
Bleeding disorders	Aneurysms, sectorial portal hypertension	Parasitic infestations
	Bleeding disorder*	Endometriosis

DU, duodenal ulcer; GU, gastric ulcer.

*In at least 25% of patients bleeding occurs from a pre-existing lesion.

Irrespective of site, GI bleeding is best addressed from a clinical standpoint in accordance with presentation of the patient:

- chronic occult blood loss: the patient is unaware of the bleeding, presents with iron-deficiency anaemia and has a positive faecal occult blood
- overt minor episodes of blood loss that prompt the patient to seek medical advice from the general practitioner: common presentation of patients with haemorrhoids, rectal tumours and inflammatory bowel disease
- acute episode of GI bleeding that causes varying degrees of hypovolaemia, in some cases life-threatening
- recurrent obscure GI bleeding.

Chronic occult gastrointestinal bleeding

Presentation with anaemia is common in patients with gastric and especially right colon cancer. Occult bleeding forms the only proven basis of screening for colorectal cancer. Using the standard haemoccult test, which has a sensitivity for cancer of around 50%, it has been shown that screen-detected tumours have a much better prognosis than tumours that present with symptoms. More importantly, randomized population-based studies have shown categorically that groups offered faecal occult blood screening have a significantly reduced mortality for colorectal cancer.

Some patients present with dyspeptic symptoms and are found to be anaemic on examination, e.g. reflux oesophagitis, peptic ulceration. Other causes of occult bleeding are much less common and usually present diagnostic problems requiring special investigations (recurrent obscure GI bleeding).

Elective presentation with episodes of gastrointestinal bleeding

Nowadays, this presentation relates most commonly to patients with colorectal disease, e.g. polyps, cancers, inflammatory bowel disease and haemorrhoids. Both cancers and inflammatory bowel disease (ulcerative colitis, Crohn's colitis, indeterminate colitis, dysentery) are usually associated with alteration in bowel habit.

Acute upper gastrointestinal bleeding

Epidemiology

The overall incidence varies widely in Western countries (40–150/100 000), with regional differences within each country. Thus, within the UK, Aberdeen has an incidence of 116/100 000 whereas Oxford, at 47/100 000, pales by comparison. Irrespective of regional differences, the epidemiology of acute upper GI bleeding has changed during the past 40 years. The highest prevalence is nowadays encountered in the elderly. This is the main reason why mortality from acute upper GI bleeding has not declined significantly despite better and earlier diagnosis by endoscopy and advances in endoscopic control of bleeding and in blood transfusion practice. The fact is that elderly people do not tolerate severe hypovolaemia and are less likely to recover if they require operative intervention than young, fit patients. Age-specific data from the UK Office for Population Census Statistics for deaths from bleeding duodenal and gastric ulcers in England and Wales demonstrate the very significant effect of increasing age, with death rates rising 400-fold between the age ranges 0–49 years and 80+ years (Figure 7.1).

Causes of acute upper gastrointestinal bleeding

The causes of upper GI bleeding in descending order of frequency are given in Table 7.3. Duodenal and gastric ulcers account for the majority. Mortality is higher in patients with a bleeding gastric ulcer, again attributable to the older age group. In some 20% of patients, the diagnosis is not clear at presentation despite urgent upper GI endoscopy. This arises because (1) there is too much blood in the stomach to permit adequate inspection, (2) the lesion is missed (e.g. Mallory–Weiss

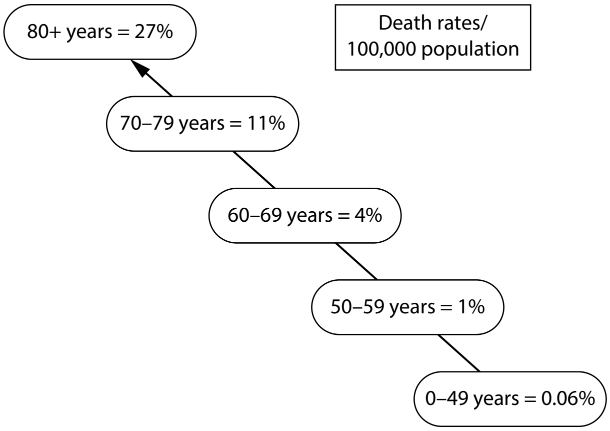


Figure 7.1 Deaths per 100 000 from bleeding duodenal and gastric ulcer in England and Wales in various age groups. Note the 400-fold increase between the age ranges 0–49 years and 80+ years.

Table 7.3 Causes of acute upper gastrointestinal bleeding: percentage distribution by disease category (average figures for the UK)

Disease category	Percentage distribution
Duodenal ulcer	29
Gastric ulcer	22
Oesophagitis and ulcer	12.0
Acute gastric lesions	7.0
Mallory–Weiss	7.0
Oesophageal varices	5.0

tear or Dieulafoy submucosal aneurysm), (3) the lesion has healed or (4) the source of bleeding is outside the stomach and proximal duodenum (upper jejunal, haemobilia, etc.).

Bleeding due to aspirin and non-aspirin non-steroidal anti-inflammatory drugs

Much of the misinformation concerning aspirin and non-aspirin non-steroidal anti-inflammatory drugs (NSAIDs) has arisen because of the poor quality of the early reports, which relied on contrast radiology for diagnosis and lacked proper controls. *All of these drugs cause drug-induced peptic ulcers in the stomach and duodenum* more commonly than acute gastric erosions, which account for only a small percentage of diagnosed cases. Aspirin increases the risks for both gastric and duodenal ulcers by two to four times, and, according to the results of a study by Faulkner *et al.* (1988), one in every 10 ulcer bleeds in patients aged 60 years or over is induced by aspirin.

The non-aspirin NSAIDs (e.g. fenbufen, benoxaprofen, indometacin, piroxicam, ibuprofen) pose a particular problem because they carry a higher risk of peptic ulceration than aspirin, and are taken largely by elderly patients who may be at increased risk of ulcer complications and death. Calculations based on published series indicate that 20% of all bleeding ulcers in patients over 60 years are caused by non-aspirin NSAIDs. A case–control study reported in 1997 demonstrated that the situation had not changed a decade later. Emergency admissions were more likely to be NSAID users than controls (31% versus 16%) and had significantly higher blood transfusion requirement, although NSAID usage did not influence mortality. On the basis of these data, it has been estimated that there are 65 000 emergency upper GI hospital admissions (perforation, bleeding, acute pain) per annum in the UK and of these 12 000 are caused by NSAID use, with 2230 deaths. A further 300 attributable deaths occur in the community (patients not admitted to hospital).

Overall, 30% of ulcer bleeds in patients aged above 60 years are due to aspirin or non-aspirin NSAIDs.

Cyclo-oxygenase inhibitors

Many agents, e.g. sucralfate, high-dose H_2 blockers and, more recently, proton pump inhibitors, have been shown to be effective in reducing duodenal lesions caused by NSAIDs, but all have been less successful in preventing gastric mucosal damage. This approach, although in established practice, is not the answer to the epidemiological problem. During the past decade there has been considerable pharmacological research aimed at the development of NSAIDs that are free of significant GI toxicity and this led to the development and clinical use of inhibitors of the cyclo-oxygenase (COX) system. COX was thought to be a single enzyme, but the subsequent research confirmed that there are two isoforms of COX: COX1 and COX2. COX1 (also known as constitutive cyclo-oxygenase) is always present and is responsible for most of the physiological prostaglandin production involved in cytoprotection, especially of the gastric mucosa. In contrast, COX2 (inducible COX) is involved in the synthesis of prostaglandins that mediate the inflammatory response (including joint inflammation). Thus, specific inhibitors of COX2 suppress the inflammatory response, while preserving the cytoprotective function of COX1. The first COX2 inhibitor to be launched in Europe was meloxicam, and this was followed by several others over a period of 10 years. The outcome of the clinical introduction of these drugs has been disappointing because their potential to decrease adverse GI effects has been more than offset by an increase in cardiovascular events, such as myocardial infarction (MI), stroke and sudden cardiovascular death in patients on long-term treatment with COX2 inhibitors. This problem was initially raised by the VIGOR study. Although this trial was designed to compare gastric toxicity between rofecoxib (COX2 inhibitor) and naproxen it demonstrated an increased risk of confirmed MI with the use of rofecoxib compared with naproxen. This was confirmed by several other clinical trials and a meta-analysis of 23 placebo- and NSAID-controlled studies with rofecoxib (including 26 000 patients), which concluded that treatment with rofecoxib was associated with an increased risk of thrombotic cardiovascular events when compared with naproxen, but not when compared with other classic NSAIDs (ibuprofen, diclofenac and nabumetone). A Prescription Event Monitoring study in the UK reported that the risk of cardiovascular events was greater in patients taking celecoxib or rofecoxib than in those taking the less selective COX2 inhibitor meloxicam. The Adenomatous Polyp Prevention on Vioxx (APPROVe) study, which was designed to evaluate the efficacy of rofecoxib in preventing recurrence of colorectal polyps in patients with a history of colorectal adenomas, was stopped prematurely because of an increased relative risk for confirmed MI during treatment with rofecoxib compared with placebo. Rofecoxib was then withdrawn from the market.

Studies have since shown that some COX2 inhibitors may tip the balance in favour of prothrombotic eicosanoids (thromboxane A_2) and thus lead to increased cardiovascular thrombotic events. COX2 inhibitors have also been shown to increase blood pressure as a result of alterations in the renin–angiotensin pathway and sodium and water retention by the kidneys. Currently, the choice of COX2 selective inhibitors for a particular patient should be selective and accompanied by information to the patients on their potential risks. Furthermore, they should be prescribed for short courses only and in the lowest effective dose.

Clinical presentation of acute upper gastrointestinal bleeding

Most patients are admitted as emergencies with haematemesis, melaena or both. Severity is dictated by the presence and extent of hypovolaemia into:

- mild: no significant hypovolaemia, includes patients who are anaemic
- moderate: hypovolaemia that responds to volume replacement (crystalloids and blood) and, thereafter, the patient is stable
- severe: active continued major bleeding rendering resuscitation with transfused blood difficult or recurrent major bleeding after successful resuscitation from the initial bleed; these are the patients at risk and include patients with bleeding oesophageal varices. In elderly patients with atherosclerosis, the hypotension may precipitate MI or a cerebrovascular accident.

Although this classification is useful as it dictates management, the category can change after initial assessment from mild to severe. Thus, complacency must be avoided and repeat clinical observation monitoring is essential even in patients with mild upper GI haemorrhage. This situation is best exemplified by the patient who develops an aortoenteric fistula after aortic replacement by a prosthetic graft. The initial (secondary) bleeds may appear trivial, yet they are warning manifestations of an impending catastrophic haemorrhage that is often fatal.

Some insist on differentiating between recurrent and persistent bleeding. The former is defined as a second episode of haematemesis or melaena associated with evidence of hypovolaemia after the initial successful resuscitation and a period of haemodynamic stability. In patients with ulcers, the endoscopic stigmata associated with increased risk of recurrence are an active spurting vessel, a visible vessel (Figure 7.2) and an adherent clot. Persistent bleeding is diagnosed when the patient requires 8 units (>60 years) or 12 units (<60 years) or more over a 48 hour period to maintain the Hb at 10 g/dL. In practice, both require measures to control the bleeding and the important decision is whether these patients should be treated endoscopically or surgically.

In addition to the severity of the bleed, the patient must be examined for stigmata of chronic liver disease that may indicate variceal haemorrhage, although these patients could equally bleed from ulcers or portal hypertensive gastropathy (see Chapter 24). Also necessary is clinical examination of the cardiovascular and respiratory systems with appropriate investigations (chest radiograph and ECG). Significant cardiorespiratory disease is an important determinant of morbidity and mortality and influences the approach used to control bleeding.

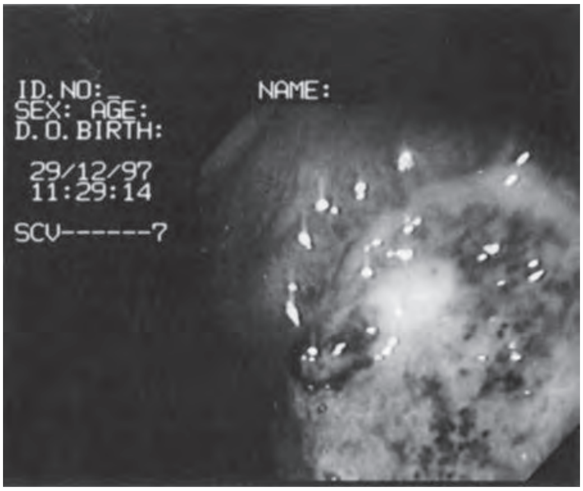


Figure 7.2 Endoscopic view of a visible vessel in a large gastric ulcer.

Management of acute upper gastrointestinal bleeding

The action plan is outlined in Figure 7.3. Diagnosis is based on upper GI endoscopy carried out within 24 hours of admission in haemodynamically stable patients and following resuscitation. The policy that all patients with acute upper GI haemorrhage should have a joint consultation by a surgeon and gastroenterologist soon after admission has much to commend it. Patients with severe continued bleeding require surgery concomitantly with volume replacement through two large-infusion cannulas.

All patients are kept on nil by mouth at least until endoscopy has been performed. The practice of insertion of a nasogastric tube is popular with surgeons but not gastroenterologists. If a tube is used, it should be of the Salem sump suction variety.

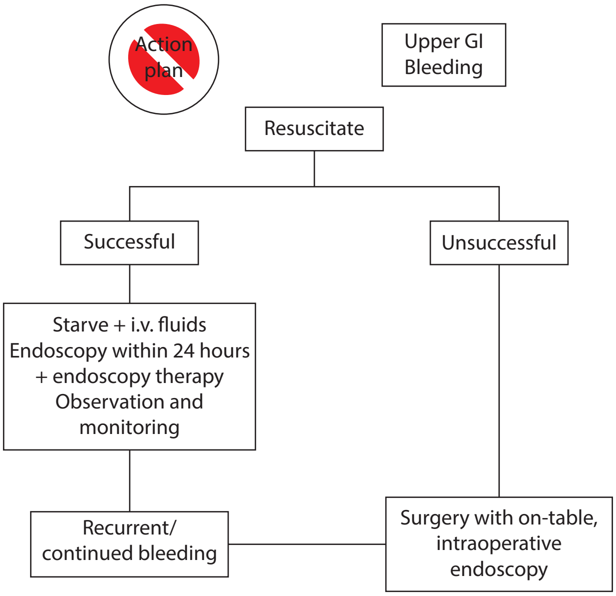


Figure 7.3 Action plan for patients with acute upper gastrointestinal bleeding.

Lavage through the tube with ice-cold saline to induce gastric mucosal vasoconstriction is practised in some North American centres but has never caught on in the UK. Without doubt, a nasogastric tube cannot be relied on to diagnose recurrent bleeding and certainly is no substitute for clinical observation and regular monitoring of pulse, blood pressure, CVP, pulse oximetry and urine output. Aside from this, nasogastric tubes may cause mucosal trauma. Thus, the case for nasogastric suction 'to keep the stomach empty', although established by surgical tradition, is by no means proven.

Pharmacological control of bleeding

The agents used are H_2 blockers, sucralfate, proton pump inhibitors, vasopressin (or analogues), somatostatin and tranexamic acid.

There has never been any material evidence that acid suppression with H_2 blockers imparts any benefit in patients with acute upper GI bleeding, although proton pump inhibitors may help. Likewise, there is no evidence that vasopressin and somatostatin are useful in ulcer bleeds. By contrast, tranexamic acid (oral or intravenous) has been confirmed by controlled, randomized studies to reduce the transfusion requirements, rebleeding rates (by 30%), need for surgical intervention (by 30–40%) and mortality (40%). The current recommended regimen consists of intravenous tranexamic acid for the first 3 days followed by oral administration for a further 3 days. To date, tranexamic acid is not widely used outside Scandinavia except as perioperative treatment for elective surgery in Jehovah's Witness patients.

Both vasopressin (and analogues) and somatostatin have been shown to be effective in the initial control of variceal bleeding (see Chapter 26).

Endoscopic treatment

Control of bleeding in the majority of patients (unless severe or catastrophic) is achieved with interventional flexible endoscopic techniques carried out by experts in a dedicated endoscopy suite with the necessary cardiovascular monitoring. Large diameter instrument channels (3.4 mm or greater) or twin-channel endoscopes (Figure 7.4) are used, as the standard endoscopes (instrument channel 2.8 mm) are very easily blocked by blood clots. The twin-channel endoscope has one channel for therapy and the other for suction. Its disadvantage lies in the large overall diameter of the instrument and consequent reduced flexibility. For this reason, some prefer the single large diameter instrument-channel endoscope and, if aspiration of luminal contents proves difficult when the therapeutic device is inserted, a nasogastric tube is passed alongside the endoscope for suction and irrigation.

The endoscopic techniques used for visible vessels and active bleeding can be classified as thermal, electrocoagulation, photocoagulation and injection therapy (Box 7.2). Sometimes, a combination of techniques, e.g. epinephrine (adrenaline) injection therapy with electrocoagulation or photocoagulation, is used and is reported to be beneficial by allowing a clearer target for the endoscopist and by reducing the heat-sink effect when thermal energy is applied.

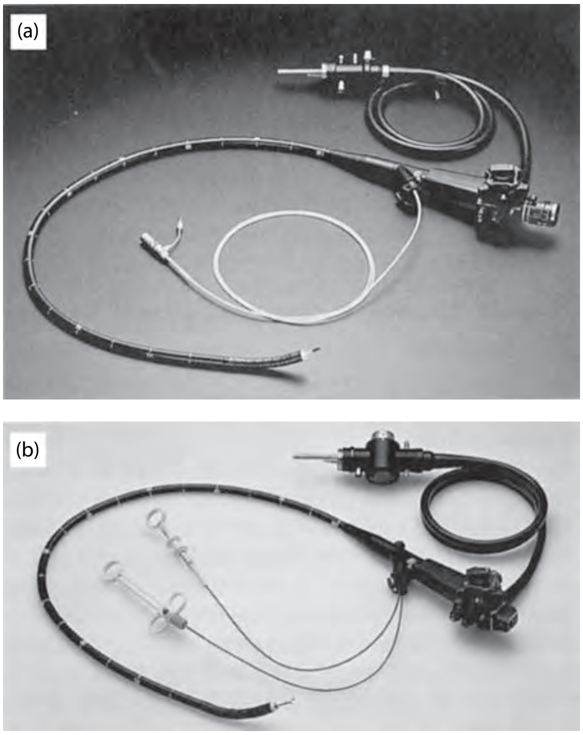


Figure 7.4 Large biopsy channel endoscopes: (a) single and (b) twin instrument channel. (Courtesy of Olympus, Japan.)

BOX 7.2 Techniques for endoscopic control of upper gastrointestinal bleeding

- Injection therapy: vasopressors, sclerosants, thrombin, mixtures
- Banding: alternative to injection sclerotherapy for oesophageal varices
- Endoscopic clipping
- High-frequency electrocoagulation: unipolar, bipolar, argon ion plasma coagulation
- Heater probe
- Photocoagulation: gas vapour lasers, diode array lasers:
  - Non-contact (laser beam)
  - Contact (with sapphire tip)

Vascular embolization

Vascular embolization is used for endoscopically or surgically inaccessible bleeding sites, e.g. liver in haemobilia, bleeding from pancreas (usually postpancreatic necrosis) and some lesions in the small intestine.

Indications for surgical treatment

With the expansion and increased efficacy of endoscopic techniques for the control of acute GI bleeding, fewer patients are treated surgically. Nonetheless, there are clear-cut clinical situations in which surgical treatment remains the patient's only hope of survival. The important determinant of survival in these seriously ill patients is a timely decision that surgical treatment is needed. Persistence with conservative management or endoscopic control until the patient's condition is seriously compromised by prolonged hypovolaemia inevitably results in demise of the patient, usually from multisystem organ failure.

The patients with acute upper GI bleeding who require surgical treatment are:

- Patients admitted with exsanguinating haemorrhage such that resuscitation by volume replacement cannot keep up with the losses. An on-table endoscopy after tracheal intubation and induction of anaesthesia is usually carried out and can be useful (e.g. detection of oesophageal varices), but often the endoscopy is unrewarding because of too much blood and clots in the stomach. Some would also include patients with an active arterial spurter in the surgical category requiring immediate surgery. Intraoperative enteroscopy (small bowel endoscopy by a long colonoscope introduced orally and guided by the surgeon through the duodenum and duodenojejunal flexure) may need to be done during the course of the operation, if the exact site of bleeding cannot be determined at laparotomy. This has replaced previous surgical manoeuvres, e.g. application of bowel clamps to different sections of the GI tract, to locate the bleeding source.
- Elderly patients (>60 years): the survival of these patients is compromised by surgical delay. Most surgeons would advise surgery if (1) more than 4 units of blood are necessary during the initial resuscitation of the hypovolaemia on admission, (2) the patient has one recurrence of bleeding after initial successful endoscopic control of bleeding and (3) there is persistent bleeding requiring 8 units of blood transfusion within 48 hours.
- Younger patients (<60 years): these usually tolerate bleeding better and therefore the threshold for surgical treatment is set higher: (1) patients requiring 8 or more units of blood during the initial resuscitation and (2) persistent bleeding requiring 12 units of blood over a 48 hour period.

Lower gastrointestinal bleeding

Epidemiology

Because rectal bleeding is common and as it rarely occasions admission to hospital, true rates are not easily ascertained. However, recent questionnaire studies indicate that about 80% of the population experience at least one episode of rectal bleeding at some stage during their lives, and, at any one point, about 7% of the population will have noticed rectal bleeding within the last 6 months. Clearly, this has very significant implications for the investigation and management of this condition.

Causes of rectal bleeding

It is useful to think of causes of rectal bleeding in terms of anatomical sites, i.e. anus, rectum, colon and upper GI tract. It is very important to remember that any lesion in the stomach, duodenum or small bowel that bleeds may present with overt rectal bleeding, if the rate of blood loss is fast enough. The causes of rectal bleeding are shown in Table 7.4.

Table 7.4 Causes of rectal bleeding by anatomical site

Anal causes	Rectal causes	Colonic causes	Upper GI causes
Haemorrhoids	Polyp	Polyp	Peptic ulceration
Fissure	Carcinoma	Carcinoma	Meckel's diverticulum (ectopic ulceration)
Perianal abscess/ fistula	Proctitis	Diverticular disease	Small bowel vascular lesion
Anal carcinoma	Rectal prolapse	Colitis (idiopathic, infective or ischaemic)	
		Angiodysplasia	

Of the common causes, haemorrhoids bleed because of trauma to the engorged vascular cushions; proctitis and colitis bleed from mucosal ulcerations; tumours bleed from dilated, fragile neoplastic, vessels; and diverticular disease from erosion of a vessel at the mouth of a diverticulum. Most overt rectal bleeding is minor in terms of amount, but, occasionally, it can be massive. When this occurs, it is usually caused by diverticular disease, angiodysplasia or an upper GI source, such as a small bowel vascular malformation or a duodenal ulcer.

Clinical presentation of lower gastrointestinal bleeding

Unlike upper GI bleeding, the vast majority of patients presenting to surgeons with rectal bleeding are seen at clinic, and only a minority are admitted as emergencies. In general, rectal bleeding can be categorized as:

- trivial
- troublesome
- suspicious
- massive.

Trivial

This type of bleeding is common and consists of occasional smears of blood seen on the toilet paper with no other symptoms. Often, there are no clinical findings, although the patient may have first- or second-degree haemorrhoids.

Troublesome

This is usually bright red, and, when caused by haemorrhoids, it typically occurs after defaecation and may drip into the toilet pan. This can be quite alarming, and, when profuse and frequent, it may lead to anaemia. Anal fissure also causes bright-red bleeding associated with defaecation, but, unlike haemorrhoids, it is usually accompanied by severe anal pain.

Suspicious

This type of bleeding should raise the clinical suspicion of colorectal cancer. Typically, rectal cancer causes quite fresh bleeding often associated with tenesmus. Left-sided colon cancer is more likely to present with dark blood mixed with or on the surface of the stool, and is frequently associated with change of bowel habit and/or lower abdominal pain. It should be noted that right-sided cancers tend not to cause overt bleeding or change of bowel habit, but rather present with iron-deficiency anaemia. This is because the faeces in the right side is still liquid, and the caecum and ascending colon are capacious and distensible. Thus, blood is well mixed with stool and obstructive symptoms are late to develop, except when the ileocaecal valve is involved.

Massive

In this case, the patient presents with profuse rectal bleeding and symptoms and signs of hypovolaemia. In many cases such bleeding stops spontaneously, but, if it does not, it constitutes a life-threatening surgical emergency. The most common cause of massive rectal bleeding is diverticular disease, but it may originate from vascular lesions of the colon or small bowel and from more proximal lesions, especially posterior duodenal ulcers. Blood acts as a cathartic, and profuse upper GI haemorrhage can pass rapidly through the small and large bowel and appear as fresh rectal bleeding. Traces of food, tablets, etc., in the blood suggest a source in the stomach or duodenum.

Management of lower gastrointestinal bleeding

Investigation

The investigation of lower GI bleeding depends on the clinical presentation (Figure 7.5). For the young patient with trivial rectal bleeding, a digital rectal examination is normally sufficient, whereas in older patients a rigid sigmoidoscopy should be carried out. With typical haemorrhoidal bleeding, a proctoscopy and rigid sigmoidoscopy should be carried out in all cases, and if no cause is found, or there is no response to treatment of haemorrhoids, a barium enema is ordered. If this is also negative, consideration should be given to flexible sigmoidoscopy and, finally, colonoscopy.

In the patient with bleeding that is suspicious of neoplasia, a thorough examination of the whole colon with barium enema and flexible sigmoidoscopy is mandatory, and, if no cause is found, with colonoscopy. Barium enema alone is not satisfactory, as the rectum is not well seen and a tortuous sigmoid colon is difficult to interpret radiologically. If polyps are seen on barium enema or sigmoidoscopy, then total colonoscopy should be arranged to visualize the whole colonic mucosa and perform polypectomy. It may be argued that colonoscopy should be carried out as a primary investigation instead of barium enema and sigmoidoscopy, but (1) it is more painful, (2) there is a higher risk of perforation and (3) in some countries the facilities and expertise for this policy are not widely available.

In massive rectal bleeding, the priority is resuscitation, but investigation must proceed promptly, usually during the resuscitation. The first step must be to exclude bleeding haemorrhoids by means of proctoscopy, and this should be followed by upper GI endoscopy. If these do not give the diagnosis and the patient remains unstable, most authorities recommend urgent mesenteric angiography, since if the bleeding is brisk enough (>0.5 mL/min) this will locate the bleeding point precisely. There is a school that favours acute colonoscopy with or without colonic lavage through a nasogastric tube, but this has been found to be impractical, and it cannot provide any degree of localization of small bowel lesions. For bleeding which is intermittent and does not show up on angiography, radionuclide-labelled red cell scanning may be used. Here, the patient's own red cells are labelled with ^{51}Cr and reinjected so that a gamma camera can be used to localize pooling of blood in the intestine. This, however, is of very limited practical value as it is necessary to carry out very frequent scans over a prolonged period.

Treatment

Medical

The only situation in which colonic bleeding can be treated effectively with drugs is an acute exacerbation of inflammatory bowel disease, the mainstay of therapy being high-dose steroids.

Endoscopic

Endoscopic treatment is used in the management of bleeding haemorrhoids in the form of sclerosant injection or banding, but it has no place in the treatment of acute colonic bleeding. There have been occasional case reports of injection therapy for bleeding diverticular disease using a colonoscope, and angiodysplasia can be treated by electrocoagulation and photocoagulation, but not usually in the acute bleeding phase.

Radiological

As with upper GI lesions, colonic bleeding can be treated by means of radiological embolization. However, more than with the small bowel, this is associated with a risk of necrosis and perforation of the bowel wall.

Surgical

The elective surgical treatment of haemorrhoids, anal fissure and colorectal causes of bleeding, such as large bowel cancer, is covered in Chapters 30 and 31. In the patient with massive bleeding from the colon or small bowel, appropriate surgical resection after attempts at preoperative localization is required. If the source is in the small bowel, it is helpful if the radiologist can leave a superselective catheter close to the bleeding site so that the surgeon can inject methylene blue down the catheter at the time of laparotomy to highlight the segment of bowel to be removed.

Occasionally, it is necessary to proceed to laparotomy without preoperative angiography, and in this case it is wise to perform intraoperative colonoscopy with on-table antegrade colonic lavage to try to localize the bleeding lesion. If this is impossible or unsuccessful, and blood appears to be confined to the colon, then colectomy and ileostomy, preserving the rectal stump for later reanastomosis, is the safest option.

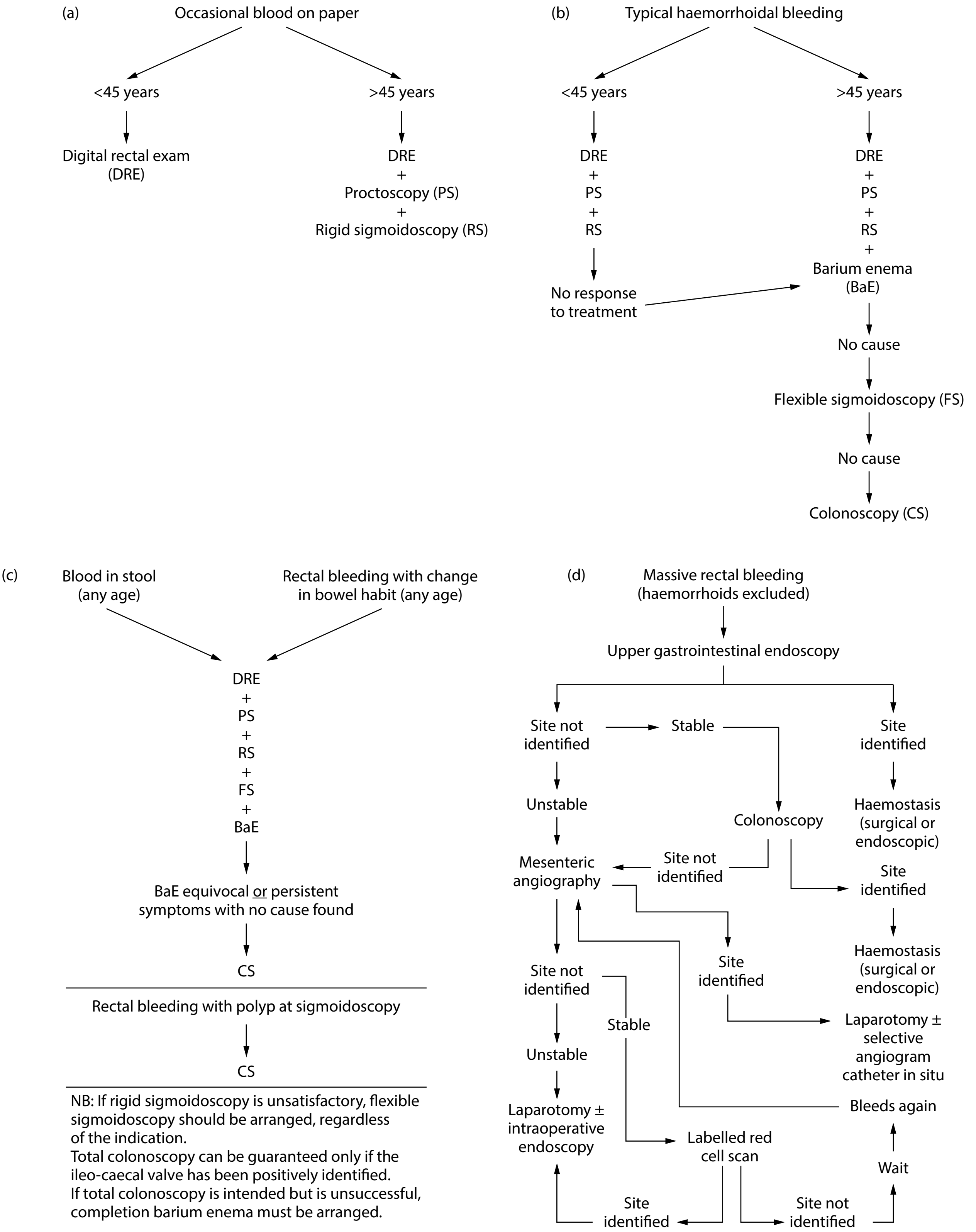
Recurrent obscure gastrointestinal bleeding

This applies to patients who experience recurrent episodes of GI bleeding (both acute and chronic) in whom no identifiable cause can be established despite full routine, often repeat, investigations. Although fortunately rare, these patients often present a management problem. The presentation may be with:

- recurrent melaena
- fresh rectal bleeding
- anaemia
- haematemesis (rare).

In the majority of these patients, the source of the recurrent blood loss is in the midgut, i.e. small bowel below the duodenojejunal junction and the proximal colon. However, some are caused by missed lesions in the foregut. This is exemplified by the Dieulafoy lesion, which is a submucosal gastric microaneurysm or unusually large artery that runs in close contact with the mucosa. In more than 80% of patients, the lesion is situated within 6.0 cm of the gastro-oesophageal junction, where it is easily missed. Haemobilia is another (albeit rare) cause of obscure recurrent GI bleeding.

Within the small bowel, the most common cause is small bowel ulceration (drug induced or associated with systemic disease, e.g. systemic lupus erythematosus, polyarthritis nodosa,


(a)
Occasional blood on paper
<45 years
>45 years
Digital rectal exam (DRE)
DRE + Proctoscopy (PS) + Rigid sigmoidoscopy (RS)
(b)
Typical haemorrhoidal bleeding
<45 years
>45 years
DRE + PS + RS
No response to treatment
DRE + PS + RS + Barium enema (BaE)
No cause
Flexible sigmoidoscopy (FS)
No cause
Colonoscopy (CS)
(c)
Blood in stool (any age)
Rectal bleeding with change in bowel habit (any age)
DRE + PS + RS + FS + BaE
BaE equivocal or persistent symptoms with no cause found
CS
Rectal bleeding with polyp at sigmoidoscopy
CS
(d)
Massive rectal bleeding (haemorrhoids excluded)
Upper gastrointestinal endoscopy
Site not identified
Stable
Site identified
Unstable
Colonoscopy
Haemostasis (surgical or endoscopic)
Mesenteric angiography
Site not identified
Site identified
Site identified
Haemostasis (surgical or endoscopic)
Site not identified
Stable
Laparotomy ± selective angiogram catheter in situ
Unstable
Bleeds again
Laparotomy ± intraoperative endoscopy
Labelled red cell scan
Wait
Site identified
Site not identified

NB: If rigid sigmoidoscopy is unsatisfactory, flexible sigmoidoscopy should be arranged, regardless of the indication.
Total colonoscopy can be guaranteed only if the ileo-caecal valve has been positively identified.
If total colonoscopy is intended but is unsuccessful, completion barium enema must be arranged.

Figure 7.5 Investigation of lower gastrointestinal bleeding.

rheumatoid arthritis, Henoch–Schönlein purpura). Small bowel tumours account for only 2–3% of cases and the lesion is then most commonly an adenocarcinoma. Peutz–Jeghers jejunal polyps and other syndromes associated with GI polyposis (see Chapter 29) often present with recurrent occult bleeding. Vascular anomalies of the small bowel can be a cause of recurrent obscure GI bleeding and are being increasingly recognized with the advent of small bowel capsule endoscopy. Telangiectasia, either as part of the hereditary condition (Rendu–Weber–Osler syndrome) or as isolated flat intramucosal telangiectasia or haemangioma, can likewise cause GI bleeding from the small bowel. One of the most common causes of acute bleeding in children is Meckel's diverticulum from ulceration of ectopic gastric mucosa. Jejunal diverticulosis can bleed and usually presents with recurrent episodes of melaena. Other causes include radiation enteritis, aortoenteric fistula (grafted patients) and endometriosis in females.

However, the most common source of recurrent obscure GI bleeding is the proximal colon and the most frequent lesion responsible is angiodysplasia. Tiny areas of angiodysplasia (about 2 mm) are easy to miss, particularly in the caecum (see Chapter 30).

Management

The primary objective is to establish the cause of the bleeding. A vast array of specialized investigations is used to elucidate the cause in the elective situation. It is important that a sensible sequence is adopted to minimize patient inconvenience and to reduce costs. In the acute bleeding situation, the investigative options of practical benefit are fewer and essentially consist of emergency angiography and on-table enteroscopy and colonoscopy (Figure 7.6).

Angiography (aortic, selective visceral arteries) plays a crucial role in these patients and is the most effective investigation. Its exact role depends on the clinical situation. In patients with active bleeding, contrast angiography seeks to demonstrate the bleeding site by documenting extravasation of contrast, whereas in the elective situation (chronic bleeding) the aim is to detect abnormal vascular patterns (Figure 7.7). The success rates of angiography in acute haemorrhage depend on bleeding rate. Thus, if aortic injection is used, the lesion has to be bleeding at the rate of at least 5.0 mL/min. With selective visceral artery cannulation, however, localization can be obtained with lower bleeding rates (0.5–2.0 mL/min).

Barium meal and follow-through is virtually useless in these patients. Much higher yields are obtained, especially for

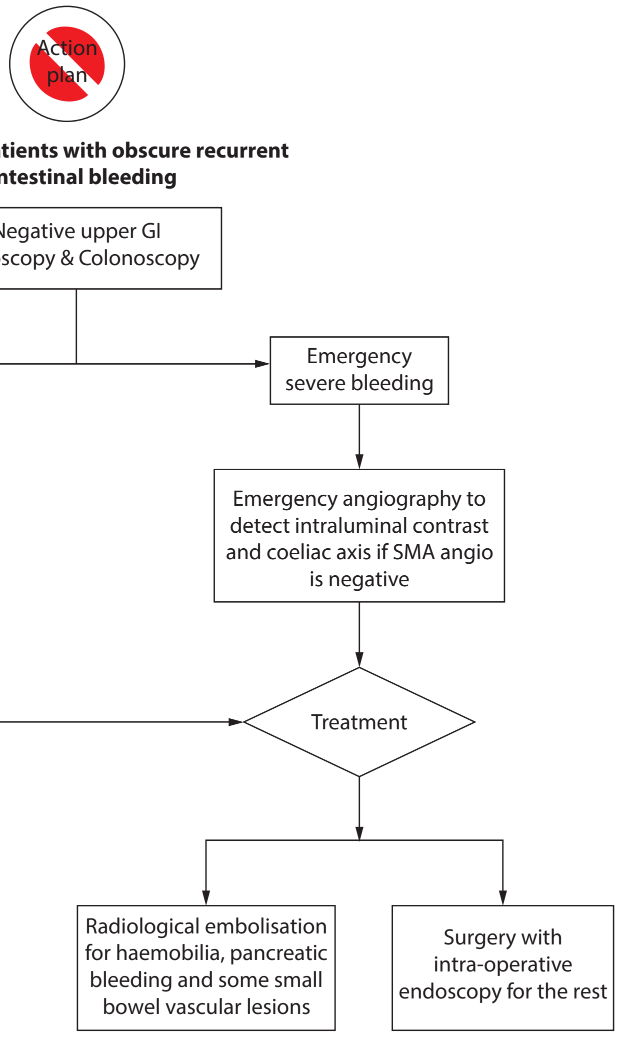

Action plan
Management of patients with obscure recurrent gastrointestinal bleeding
Negative upper GI Endoscopy & Colonoscopy
Elective situation
Emergency severe bleeding
1. Small bowel enteroclysis
2. Enteroscopy
3. Radionuclide studies: (^{99m}Tc-pertechnetate for Meckel's ^{99m}Tc-SC, labelled RBCs)
4. ERCP (haemobilia), followed by coeliac axis angiography if ERCP is positive
5. Angiography to detect vascular lesion in gut
Emergency angiography to detect intraluminal contrast and coeliac axis if SMA angio is negative
Treatment
Radiological embolisation for haemobilia, pancreatic bleeding and some small bowel vascular lesions
Surgery with intra-operative endoscopy for the rest

Figure 7.6 Action plan for recurrent obscure gastrointestinal bleeding.

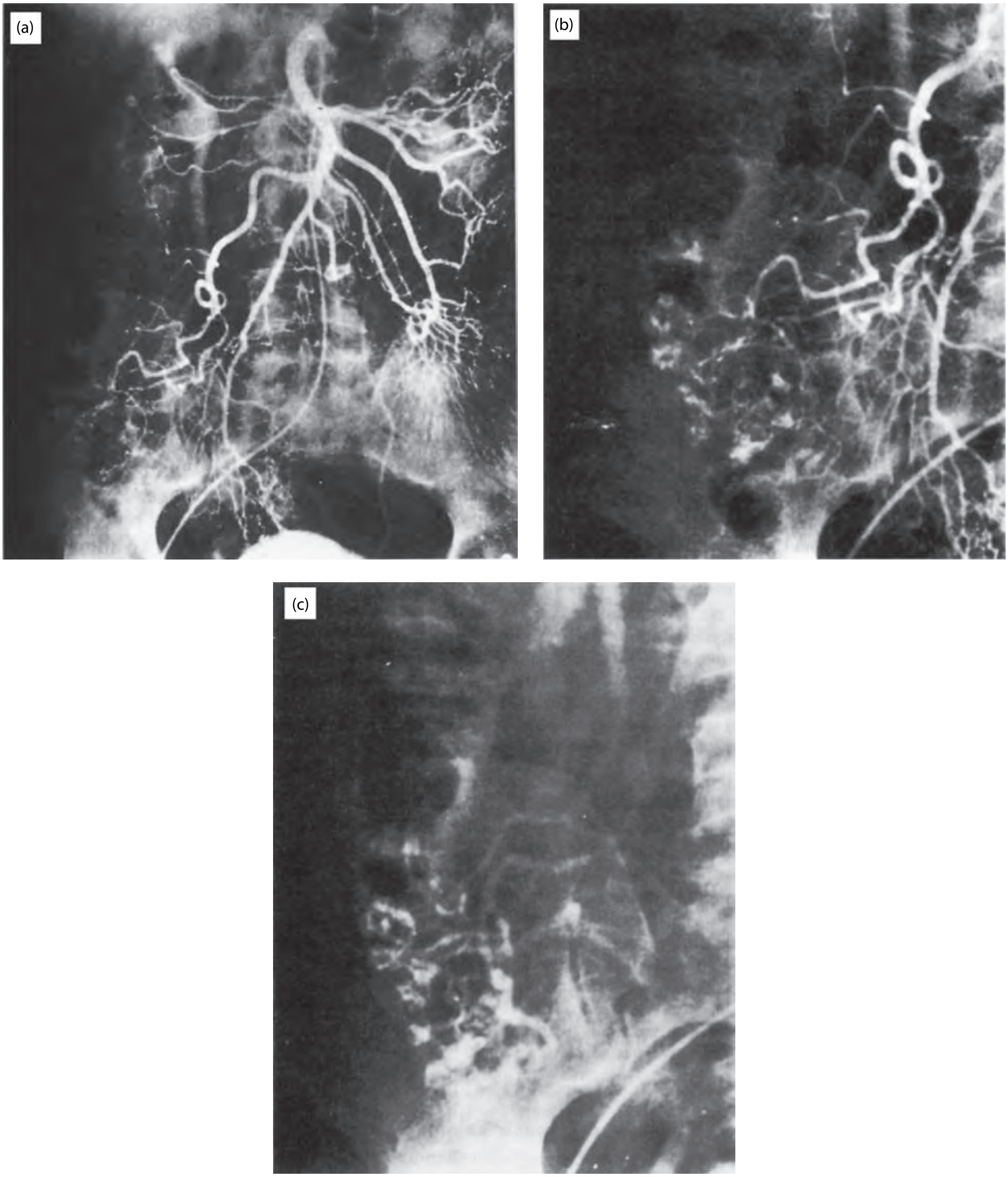


Figure 7.7 Selective superior mesenteric angiogram in chronic bleeding demonstrating vascular abnormality in the caecum. The angiographic features are: (a) prominent and early filling of clusters of irregularly dilated small vessels with the caecal wall supplied by the ileocolic artery; (b) persistence of contrast within the clusters of small blood vessels; and (c) early and profuse filling of enlarged draining veins from the affected area.

the detection of small bowel tumours, ulcers and strictures, by enteroclysis (small bowel enema) with compression fluororadiology. Radionuclide scanning is used for three purposes: (1) to quantify blood loss in the chronic situation (5 day test using ^{51}Cr-labelled autologous RBCs); (2) for localizing blood loss by technetium-99m (^{99m}Tc)-sulphocolloid in the first instance, followed by labelled autologous RBCs; and (3) for the diagnosis of bleeding Meckel's diverticulum, when ^{99m}Tc-pertechnetate is used to display the ectopic gastric mucosa (not the ulcer or the bleeding point itself). Endoscopic retrograde cholangiopancreatography is useful in cases of suspected haemobilia or if the bleeding is likely to be of pancreatic origin.

Ultimately, the treatment depends on the nature of the lesion, its surgical accessibility, the severity of the clinical situation and the condition of the patient. In general, all patients with life-threatening bleeding from the gut itself require laparotomy. Those in whom bleeding is from liver and pancreas are best managed with radiological embolization in the first instance.

Retroperitoneal haemorrhage

Retroperitoneal haemorrhage is always caused by a serious condition and, aside from hypovolaemia, which may be extreme and life threatening, it is also often accompanied by adynamic ileus. Retroperitoneal haematoma/bleeding is caused by:

- contained rupture of an abdominal aortic aneurysm (commonest)
- severe haemorrhagic pancreatitis
- severe pancreaticoduodenal trauma
- warfarin therapy
- haemophilia and other coagulopathies
- renal (renal cell carcinoma and angiomyolipoma) and adrenal tumours (rare)
- polyarteritis nodosa (rare)
- nephritis (rare)
- segmental arterial mediolysis (very rare)
- primary amyloidosis (very rare).

Common clinical settings include patients with leaking abdominal aneurysms and victims of retroperitoneal trauma involving pancreas, duodenum and kidney (see Chapter 14). Retroperitoneal bleeding, which is particularly difficult to control, is encountered in some patients as a complication of severe necrotizing pancreatitis see Chapter 27). This combination is often fatal. An iatrogenic cause of retroperitoneal bleeding that has been responsible for some fatalities is aortoiliac trauma inflicted most commonly by the Veress needle during the induction of a closed pneumoperitoneum, although some of these injuries have been caused by the first (optical) trocar/cannula. These mishaps are entirely preventable by adoption of the proper technique, use of open laparoscopy or visually guided insertion systems.

Treatment

Patients with retroperitoneal haematoma/bleeding are managed in high-dependency/intensive care units because of their unstable condition and need for monitoring and resuscitation, including blood transfusions. Patients with coagulopathy or inherited clotting disorders require appropriate clotting factor replacement therapy and, in these patients, conservative treatment is indicated. Patients bleeding as a result of warfarin therapy are treated with an infusion of prothrombin complex concentrates which contain all the deficient factors (II, VII, IX and X). Emergency surgery is required for ruptured aortic aneurysm and ruptured retroperitoneal tumours.

Spontaneous rupture of the kidney affects either the collecting system or parenchyma and, in most cases, the non-traumatic rupture is associated with underlying renal disease. Rupture of the collecting system can be managed satisfactorily by drainage (percutaneous nephrostomy or ureteral tube drainage). Rupture of the renal parenchyma accompanied by severe renal haemorrhage requires emergency surgical exploration of the kidney and nephrectomy.

Transfusion of blood and plasma products

Rather than the administration of whole blood, modern blood transfusion is based on the selective intravenous infusion of the required component of the blood that is necessary in a particular patient. This practice ensures safe, economic use and increases the therapeutic scope. Thus, a single donation may be used for treating a variety of disorders.

Essentials of safe blood transfusion practice

Modern safe blood transfusion practice is based on:

- careful selection and testing of donors
- quality assurance testing of blood and blood products prepared and issued by the transfusion centres
- blood grouping, antibody screening and cross-matching
- standardized blood-ordering policies
- strict guidelines for administration and monitoring of patients receiving blood products by the clinicians
- local and central reporting of adverse effects when encountered.

Two blood samples (clotted and anticoagulated in EDTA) are sent with the completed request form to the blood transfusion department once the decision is taken that the patient needs a transfusion. The patient's red cells are grouped for ABO and rhesus (Rh) (D). The antibody screen is carried out on all serum samples to detect the presence in the recipient of any clinically important antibody that would haemolyse the transfused cells. Cross-matching entails testing the red cells from the donor units against the patient's serum. This constitutes the final safeguard against incompatibility. In some countries cross-matching is considered unnecessary if detailed blood group and screening have been performed, but within the UK it remains, by general consensus, standard practice.

The final step in the safe administration of blood is the check by two people that the information on the transfusion request form, the labels on the unit of blood and the patient identification (from case notes and wrist band) all agree. Regrettably, this remains the weakest link in the safety chain and most deaths from blood transfusion in the UK are tragic consequences of errors in identification.

General and medicolegal aspects

Different regulations concerning the transfusion of blood and plasma products apply within various countries. In the UK, plasma products fall within the scope of the UK Medicines Act (1968) and the associated European Community Legislation. All such plasma products have thus to be licensed and thereafter are considered as prescription-only medicines (POM; i.e. can only be prescribed by a fully qualified medical practitioner); each must be accompanied by an approved summary of product characteristics.

In contrast, whole blood and blood components are exempt from the UK Medicines Act (1968) and are therefore supplied by the regional blood transfusion centres as unlicensed products. However, blood components are considered in clinical practice as POM. It is a medicolegal requirement within the UK that details of any blood component infusion are fully documented in the patient's case notes: *indication, date/time of issue, donation number, ordering physician, duration of infusion, who gave the infusion, who checked it prior to administration and time the infusion was commenced. The written procedures for infusing blood and blood components are issued by the hospital blood transfusion department. The hospital management is responsible for ensuring that the written procedures are updated and made available to all staff (medical and nursing) who may have to administer transfusions.*

In the UK, all blood and plasma products are derived from voluntary, non-remunerated donors aged 18–65 years. In accordance with the policy of the Department of Health, donors are carefully selected and all donations are tested for known markers of disease. Currently, these are syphilis, human immunodeficiency virus (HIV) 1, HIV-2, hepatitis B and hepatitis C. Sterilization of other transmitting agents not detected by donor screening is not guaranteed. Thus, the risk of disease transmission, although very low, cannot be ignored. Pertinent examples of this include the transmission of acquired immunodeficiency syndrome (AIDS) to haemophilic patients and their partners in the recent past. The current concern is with the recently identified hepatitis G virus.

Plasma products

Plasma products are produced at protein fractionation centres from pools derived in the UK from voluntary, non-remunerated donors. These products (Table 7.5) do not require ABO compatibility with the patients and are used in clinical practice for the following purposes:

- As coagulation factors for specific deficiency states, either to stop spontaneous bleeding or to cover an operation/intervention in a patient with such a deficiency, e.g. haemophilia A (congenital factor VIII deficiency).
- To provide passive immunity to non-immunized individuals exposed to a serious infective agent (viral or bacterial), e.g. human tetanus immunoglobulin.
- In the prophylaxis of haemolytic disease of the newborn due to Rh incompatibility and sensitization of Rh (D)-negative women. Human anti-D immunoglobulin is also used in the management of Rh (D)-incompatible blood transfusion following the accidental transfusion of Rh (D)-negative individuals with Rh (D)-positive blood.
- In the management of autoimmune thrombocytopenic purpura (AITP). High-dose intravenous human immunoglobulin can produce remissions of varying duration. IgG is also used to treat acute haemorrhage from this condition when conventional therapies have failed and to cover patients with idiopathic thrombocytopenic purpura requiring surgery including elective splenectomy.
- As volume replacement fluid (ALBA 4.5%), as a plasma volume expander and in short-term management of hypoproteinaemic patients (human albumin solution 20%).

Intravascular (serum) albumin accounts for 40–45% of the total body albumin. Under normal conditions, albumin has a half-life of 15–20 days. The normal level is achieved by a balance between synthesis (in the liver) and breakdown (predominantly intracellularly by lysosomal proteases) through a feedback regulation controlled by the oncotic pressure of the hepatic extravascular space. Serum albumin is a key protein in circulatory and transport physiology. In the first instance, it accounts for more than 50% of the total plasma proteins and, for this reason, it is the principal agent responsible for the plasma oncotic pressure essential for the exchange of fluid (and thus oxygen, nutrients and waste products) between the microcirculation and the interstitial fluid space. Albumin, through its oncotic action, ensures the re-entry of fluid from the interstitial space to the capillary bed and, in this fashion, it stabilizes the CBV. Hypoalbuminaemia from any cause thus results in fluid sequestration in the interstitial space and serous cavities (oedema/ascites) with contraction of the effective circulatory blood volume. Serum albumin also provides an important transport service as a carrier for hormones, enzymes, drugs, metabolites (e.g. bilirubin) and toxins. The combination with albumin renders all of these substances water soluble and inactive.

ALBA 4.5% is iso-osmolar and is supplied in bottles of 400 mL containing approximately 45 g/L plasma protein, of which >95% is albumin. It is indicated in acute blood volume loss (instead of or with crystalloids), and in the treatment of burns after the first 24 hours (when crystalloids are preferred) to maintain a plasma albumin close to 25 g/L. This is necessary to ensure a colloid osmotic pressure above 20 mmHg (2.7 kPa) and thus the return of fluid to the circulation at the venous end of the capillaries. ALBA 4.5% is also useful in surgical practice in conditions characterized by severe acute albumin losses, e.g. acute pancreatitis and small bowel infarction. Rapid infusion of ALBA 4.5% can cause vascular overload with pulmonary oedema in patients with diminished cardiorespiratory reserve or hypertension. Thus, administration in these patients should be cautious and always monitored by a CVP line. Less than 10% of the infused albumin leaves the intravascular compartment during the first 2 hours after infusion. Thus, the CBV will increase from 1 to 3 hours after administration and CVP monitoring in these patients should be continued for at least 4 hours after the albumin infusion has been stopped.

Human albumin solution 20% is an aqueous solution containing 20% w/v protein in 100 mL with albumin forming 95% or more of the protein. The solution is hyperosmolar, i.e. it acts as oncotic agent, drawing fluid from the extravascular compartment to the intravascular compartment. Hence, there is a real danger of overload in susceptible patients, e.g. patients with diminished cardiorespiratory reserve and hypertensive patients. For this reason, human albumin solution 20% is usually contraindicated in patients with severe anaemia or cardiac failure. Human albumin solution 20% is indicated in the short-term management of severely hypoproteinaemic patients with diffuse oedema that is resistant to diuretic therapy (e.g. nephrotic syndrome) and in the initial management of patients with leakage due to damage of major lymph trunks in

Table 7.5 Plasma products used in clinical practice

Plasma product	Indication	Route of administration	Shelf life	Storage
Coagulation factors				
Factor VIII concentrate	Haemophilia A	i.v.	24 months	Light-protected storage at +2–8°C
Factor IX concentrate	Single/multiple congenital deficiency of factor IX, II or X Single/multiple acquired prothrombin complex factor deficiency (including reversal of Coumarin anticoagulant therapy)	i.v.	36 months	Light-protected storage at +2–8°C
Immunoglobulin products				
Human immunoglobulin	X-linked primary hypogammaglobulinaemia Late-onset hypogammaglobulinaemia Hypogammaglobulinaemia secondary to haematological malignancies Children with symptomatic HIV infection who have recurrent bacterial infections Idiopathic thrombocytopenic purpura Kawasaki disease In allogeneic bone marrow transplant to prevent infections and graft-versus-host disease	i.v.	Lyophilized product = 24 months. Used immediately after reconstitution	Light-protected temperature not exceeding 25°C
Human normal immunoglobulin (250 mg protein per vial, 90%+ = IgG)	Hepatitis A prophylaxis Prevention of measles in immunocompromised patients Replacement therapy in primary immunodeficiency syndromes	Deep i.m.	36 months	Light-protected storage at +2–8°C
Human hepatitis B immunoglobulin (1000 IU per vial)	Prophylaxis of hepatitis B after parenteral exposure or contamination of a wound/conjunctiva by blood/secretions containing HBsAg, recent sexual contacts of diagnosed hepatitis B cases Newborn babies whose mother has had acute hepatitis B in the last trimester or is a known HBsAg-positive carrier	Deep i.m.	36 months	Light-protected storage at +2–8°C
Human tetanus immunoglobulin (250 IU per vial)	Used in conjunction with tetanus toxoid in patients with 'at-risk wounds' of tetanus and who have not been adequately immunized against tetanus or whose immune status is in doubt	Deep i.m.	36 months	Light-protected storage at +2–8°C
Human anti-D immunoglobulin (various strengths)	Rh (D) sensitization in Rh (D)-negative mothers and Rh (D)-incompatible blood transfusion	Deep i.m.	24 months	Light-protected storage at +2–8°C
Human varicella zoster immunoglobulin (250 mg HVZ immunoglobulin per vial)	Prophylaxis of immune compromised individuals/patients exposed to varicella or zoster	Deep i.m.	36 months	Light-protected storage at +2–8°C
Human rabies immunoglobulin (800 IU per vial)	Used in conjunction with rabies vaccine in individuals exposed to rabies and who are not considered adequately immunized	Deep i.m.		Light-protected storage at +2–8°C
Albumin products				
Human albumin solution 4.5%	Acute blood volume loss. Severe burns after first 24 hours. Therapeutic plasmapheresis Severe acute albumin losses	i.v.	36 months	Light-protected storage below 25°C
Human albumin solution (20%)	Short-term management of hypoproteinaemic patients Chronic liver disease with refractory ascites and peripheral oedema Severe burns requiring plasma volume expansion	i.v.	18 months	Light-protected storage below 25°C

HBsAg, hepatitis B surface antigen; HVZ, human varicella virus; Rh, rhesus.

the chest or peritoneal cavities (chylous ascites, chylothorax) following surgical interventions (usually radical resections for oesophageal and gastric cancer). In this instance, human albumin solution 20% is administered to maintain the serum albumin above 25 g/L until surgical treatment designed to secure the intrathoracic or intra-abdominal lymph leak is carried out as a matter of some urgency because of the daily profound losses. Surgical control is much preferable to peritoneovenous shunting for chylous ascites, as it constitutes definitive therapy and prevents the risks of peritoneovenous shunting, which include episodes of DIC and infection. Human albumin solution 20% is also used in patients with hypoproteinaemia associated with refractory ascites and peripheral oedema due to chronic liver disease, and in the clinical management of patients with severe burns in whom plasma volume expansion is required.

Other plasma products

Other plasma products are listed in Box 7.3. The important ones, from a surgical standpoint, are fibrin glue, intravenous human immunoglobulin, tetanus immunoglobulin and intravenous hepatitis B immunoglobulin.

Human fibrin glue

There are a number of proprietary preparations of this glue based on a similar formulation (Tisseal-Immuno; Fibrin Sealant Kit-SNBT). The two solutions (thrombin and fibrinogen) are derived from pooled non-remunerated donations and heat inactivated against all known human viruses. The two solutions have to be reconstituted immediately prior to use and special delivery sets for both open and laparoscopic surgery are needed for application to the bleeding surface or suture line. Essentially, the delivery sets consist of twin syringes (one for each of the two solutions) with a Y-connection to a twin barrel segment, with the two barrels joining at the end to allow mixing of the two solutions and thus the polymerization of fibrinogen to fibrin (see also Chapter 2).

Human immunoglobulin for intravenous use

Human immunoglobulin for intravenous use (IVIgG) can be used in the following:

- patients with AITP to cover surgery and childbirth
- antibody-deficient states, e.g. primary and secondary hypogammaglobulinaemia (due to myeloma and chronic lymphatic leukaemia and children with AIDS) to prevent recurrent bacterial infections.

IVIgG is not a substitute for orthodox management of AITP including steroids and splenectomy (when steroid therapy fails). It is used in the management of thrombocytopenic bleeding in which high-dose therapy (1–2 g/kg over 1–5 days) can produce a significant rise in the platelet count, but this is of a variable duration. When a patient with AITP requires surgery (including splenectomy), IVIgG is administered preoperatively if the platelet count is very low. High-dose IVIgG can cause renal failure, especially in the elderly. Thus, it is administered cautiously and only when strictly necessary in the elderly and in patients with pre-existing renal disease.

Blood components

Whole blood and blood components are prepared in regional transfusion centres from individual donations or pools of a small number of donations. The ABO and Rh (D) of each unit of whole blood and cell preparations are stated on the pack label of each unit and these products can only be administered after cross-matching to ensure compatibility with the patient's serum, except in dire life-threatening emergencies, when non-cross-matched group O Rh (D)-negative blood may be given. The ABO group of platelet concentrates is usually compatible with that of the patient but, although highly desirable, this is not essential. Platelets from Rh (D)-positive donors may produce anti-Rh (D) antibodies when transfused to Rh (D)-negative patients (important in a female of childbearing age). The ABO group of FFP should be compatible with that of the patient. The ABO compatibility of cryoprecipitate with that of the patient is desirable, although not essential.

BOX 7.3 Other plasma products

- High-purity factor IX concentrate
- Human immunoglobulin (for intravenous use) 10 g dose size
- Fibrin glues
- Human fibrinogen concentrate
- Human thrombin
- Intravenous tetanus immunoglobulin
- Intravenous cytomegalovirus immunoglobulin
- Intravenous hepatitis B immunoglobulin

Ordering of blood

Emergency transfusion

The decision to administer an emergency blood transfusion (i.e. administering blood that has not been fully cross-matched) is entirely the responsibility of the senior clinician concerned in the management of the patient. In the UK, NHS hospitals keep a small stock of emergency blood (O negative, Kell negative) in the blood bank refrigerator for this purpose. Each emergency unit has a blood bank form attached that should be completed as soon as possible and returned to the blood bank. When such units are used, the blood transfusion staff must be informed immediately. Extra care should be taken to monitor the patient closely for at least the first 5–10 minutes of infusion to detect any evidence of acute reactions. Unused units should be returned to the blood bank refrigerator within 20 minutes of collection if not required. If blood is transferred with a patient to another unit the blood transfusion department should be informed so that a transfer docket can be prepared for the receiving hospital.

Elective situation

Unnecessary ordering of cross-matched blood constitutes bad practice and is wasteful of a limited resource. Many operations do not usually require a perioperative blood transfusion, others may do and some incur a need for red cells as a matter of routine. Nowadays, blood for elective surgery is ordered in accordance with the 'blood tariff' policy. This is based on the average requirement of a particular operation (Table 7.6).

Type and screen (group and hold, group and save) is all that is required for all operations that do not usually require blood. The patient's ABO and Rh (D) type are determined and the serum is screened for IgG antibodies against red cells. The patient's serum is kept in the blood transfusion laboratory for 7 days. Should blood be required, cross-matched red cell units can be available in 15 minutes.

Predeposit

The benefit of this approach (autologous as distinct from allogeneic transfusion) is the elimination of disease transmission and allergic and incompatible reactions (other than procedural errors). There are established UK guidelines for autologous blood transfusion. Patient selection includes establishing the fitness of

Table 7.6 Blood tariff for elective surgery

Operation	**Blood requirements**		
	Routine tariff group and screening(G + S) procedure, no units cross-matched	**Increased tariff considerations**	**Indication leading to increased tariff**
General surgery			
Abdominoperineal resection	4		
Bowel resection	3		
Breast biopsy/lumpectomy	G + S		
Cholecystectomy	G + S		
Partial gastrectomy	G + S		
Total gastrectomy	3	4	Thoracotomy
Haemorrhoidectomy	G + S		
Hernia repair	G + S		
Ileostomy	G + S		
Laparotomy	G + S	2	Malignancy
Liver biopsy	G + S		
Radical mastectomy	2		
Simple mastectomy	G + S		
Splenectomy (elective)	2		
Thyroidectomy	G + S		
Vagotomy (truncal, RSV)	G + S		
Varicose vein operations	G + S		
Urology			
Partial cystectomy	2		
Total cystectomy	4		
Transurethral resection of bladder lesions	G + S	2	Larger tumours
Nephrectomy	2		
Nephrolithotomy	2		
Prostatectomy	2		
Thoracic surgery			
Oesophagogastrectomy	4		
Hiatus hernia	G + S		
Pneumothorax	G + S		
Thoracotomy for pulmonary resection	3	4	Reoperation
Mediastinoscopy	2		
Arterial surgery			
Aortic aneurysm	6	10–12	Ruptured
Femoropopliteal bypass	3		

After Napier *et al. BMJ* 1985;**291**:799–801, courtesy of the editor.

the patient for the procedure. Predeposit is arranged through the blood transfusion department and between 2 and 4 units can be harvested preoperatively. The blood is collected, labelled and stored as in allogeneic (donor) voluntary donations, and the autologous units ABO and Rh (D) are grouped and cross-matched with the patient's serum. Despite the initial enthusiasm, the experience with predeposit has been disappointing. It is more expensive that allogeneic blood transfusion. The blood is wasted if the patient does not need perioperative transfusion as it cannot be administered to other patients, and if the patient requires blood in excess of the predeposited units, all possible advantage is lost. Currently, autologous blood transfusion seems a sensible option in fit patients (aged 18–65 years, free from cardiovascular, cerebrovascular and respiratory disease, no sepsis, malignancy or haematological abnormality) requiring an operation that is very likely to require modest amounts of perioperative blood transfusion and who live near the hospital.

Controlled haemodilution

In this procedure, the patient is bled immediately before surgery (after induction of anaesthesia) down to a haematocrit of 25–30% with CVP monitoring while the blood volume is maintained by infusion of crystalloids. The collected autologous blood is transfused after the operation. Aside from reducing the need for donor blood, controlled haemodilution by preoperative isovolaemic bleeding has a number of advantages, which emanate from the reduction in the blood viscosity: increased capillary perfusion, and lower incidence of deep vein thrombosis and postoperative renal failure. Controlled haemodilution is used extensively in cardiothoracic surgery and for patients undergoing major orthopaedic operations. It is practised less frequently in general surgery. Hb-based blood substitutes are being used for controlled 'balanced' haemodilution (see Blood substitutes).

Intraoperative blood salvage

This includes automated blood salvage by the cell-saver equipment (Haemonetics) and more simple manual systems for storage and reinfusion of red cells, as exemplified by the Solcotrans autologous collection system. Both techniques are only applicable to clean operative sites without bacterial, bowel or tumour cell contamination. The Haemonetics cell-saver system is a completely automated device that aspirates, anticoagulates and filters the extravasated blood from the operative field. The red cells are then washed before being transfused in a PCV of 0.50. It is used extensively during liver transplantation and in Jehovah's Witness patients undergoing major surgery.

The manual techniques have limited capacity (a few units), such that their usefulness is suspect, and it has been argued that in many instances they are employed under circumstances in which blood transfusion was probably unnecessary.

Transfusion of fresh blood

The transfusion of fresh blood for bleeding disorders has been largely replaced by FFP and therapy with other blood components, depending on the nature of the deficiency as determined by the coagulation screen.

Whole blood

Whole blood has a haematocrit of 0.35–0.45. The cells are suspended in 300 mL of a 4:1 mixture of donor plasma and a nutrient citrate anticoagulant solution. With storage (maximum 35 days at 2–6°C), changes occur in both the composition of the supernatant solution and the haematocrit. Thus, at expiry the K^+ level averages 9.0 mmol (1.0 mmol for fresh blood), the lactate 10.0 mmol (1.0 mmol for fresh blood) and the hydrogen ion activity 55.0 nmol (20 nmol per unit of blood for fresh blood). Thus, with storage, whole blood becomes increasingly acidotic. The deterioration is much more rapid when whole blood is kept out of the refrigerator or blood transport box. If the unit of whole blood has been left out at room temperature for more than 30 minutes prior to commencement of infusion, the unit should be discarded. The transfusion of each unit of blood should be complete within 4 hours of spiking the pack. A blood warmer is used when large volumes of whole blood are transfused over a short time and when the patient's plasma contains cold-acting antibodies. Whole blood is not suitable for the correction of symptomatic or critical anaemia.

Red cell preparations

As with whole blood, each unit of red cells is obtained from a single donor and consists of RBCs suspended in a total volume of 300 mL. There are minor differences in the red cell preparations from various regional transfusion centres. The Scottish National Blood Transfusion Service supplies two preparations: red cells and red cells in additive solution. A unit of red cells has a haematocrit of 0.65–0.75 with 80 mL of a 4:1 mixture of donor plasma and a nutrient citrate anticoagulation solution. A unit of red cells in additive solution has a haematocrit of 0.55–0.65 with 100 mL of storage medium and only a small residue of anticoagulated plasma. Both preparations have a much lower Na^+ (15–20 mmol) and hydrogen ion activity (2–5 nmol) than whole blood. They are used primarily to increase tissue oxygenation after haemorrhage and in patients with anaemia. They are both stored at 2–6°C and their maximum shelf-life at this temperature is 35 days. The same precautions on safe usage apply as with transfusion of whole blood.

Platelets

Platelet preparations are available as concentrates from single or multiple donors, with each unit containing 55×10^9 platelets suspended in 50–69 mL of a 4:1 mixture of donor plasma and nutrient anticoagulant solution. They can also be prepared from a single donor by an apheresis procedure (platelet apheresis). Each unit of the latter contains at least 240×10^9 platelets suspended in 120–250 mL of a 4:1 mixture of donor plasma and a nutrient anticoagulant solution. Patients with platelet alloimmunization (usually because of previous sensitization) can only receive platelet apheresis units and these have to be human leucocyte antigen (HLA) compatible or be cross-matched and shown to be compatible with the patient's serum. Patients with profound cellular immunodeficiency or those receiving transfusions from related donors should be transfused with irradiated platelets.

Infusion of platelets is indicated (usually prophylactically) in thrombocytopenia or in the presence of a functional abnormality of the platelets (e.g. Glanzman disease) when these are associated with bleeding problems. In surgical practice, platelet cover is needed when the count falls below 50 000. In AITP, human IgG is administered to raise the platelet count before platelet transfusion is considered. The indications for platelet transfusion are summarized in Box 7.4.

The development of antibodies to HLA class I antigens occurs in approximately 50% of patients who receive intensive blood product support. These antibodies cause febrile reactions, reduce successful take of a bone marrow transplant in patients with aplastic anaemia, and can also lead to platelet refractoriness. This is diagnosed if the patient's platelet count fails to rise by at least 10×10^9 on the day after a platelet transfusion. Platelet

BOX 7.4 Indications for platelet transfusion

- Surgical
  - Bleeding and thrombocytopenia
  - Cover for operative interventions: platelet count: <50 000, platelet dysfunction
  - Acute DIC (with FFP)
  - Massive blood transfusion
  - Postcardiac bypass thrombocytopenia
- Medical
  - Marrow-suppressed patients (intensive care)
  - Aplastic/hypoplastic anaemia
  - Disorders of platelet function: Glanzmann disease, Bernard–Soulier syndrome, platelet storage pool deficiency

refractoriness due to alloimmunization can be prevented to some extent by leucocyte-depleted platelet transfusion (by either plateletpheresis or the use of filters to remove the leucocytes).

Fresh frozen plasma

Each unit is obtained from a single donation and consists of 200–300 mL of plasma with 40–60 mL of citrate anticoagulant nutrient mixture. The plasma proteins in FFP are in similar concentrations to those in normal plasma. FFP is stored at –30°C and with this frozen storage has a shelf-life of 12 months. It is thawed rapidly at 37°C immediately before use and infused intravenously through a standard blood component transfusion set containing an on-line filter. The ABO group of the FFP must be compatible with that of the recipient. The infusion of 1 unit of FFP should be completed within 4 hours of thawing.

FFP is used in the following clinical situations:

- to correct isolated plasma protein deficiencies, e.g. isolated deficiencies of factors II, V, VII, X, XI, XIII, pseudocholinesterase, antithrombin III and C1 esterase inhibitor
- to reverse oral anticoagulation with warfarin/coumarin compounds if prothrombin complex concentrate is not available; normally, reversal of anticoagulant is indicated in the presence of bleeding
- in patients with liver disease, major hepatic resections and severe liver injuries to provide haemostatic support and to cover operations and interventions
- in the treatment of patients with DIC to replace factors consumed by the pathological process
- in patients who develop a bleeding diathesis after large-volume blood transfusion; in this situation, however, platelet transfusion is more commonly needed first
- in the treatment of thrombotic thrombocytopenic purpura, in which FFP is usually combined with plasma exchange.

Cryoprecipitate

Cryoprecipitate is a concentrate of factor VIII, von Willebrand's factor and fibrinogen, and is obtained by freezing and thawing plasma. One unit of cryoprecipitate consists of 20 mL containing 150–300 mg of fibrinogen and 80–120 IU of factor VIII. It is stored at –30°C and at this temperature has a shelf-life of 12 months. It is thawed rapidly at 37°C immediately before use and infusion should be complete within 4 hours of thawing.

Cryoprecipitate is used for the following conditions:

- fibrinogen deficiency and dysfibrinogenaemia
- uraemic bleeding
- von Willebrand's disease.

Special blood components

These include irradiated components, frozen, thawed and washed red cells, components for neonatal use, leucocyte-depleted cellular components, CMV-negative cellular components, apheresis plasma and cryosupernatant. Washed red cells are used in patients who develop non-haemolytic transfusion reactions. Frozen, thawed and washed red cells are indicated in patients who have rare antibodies.

As infection with cytomegalovirus (CMV) can be lethal in patients following bone marrow transplantation, leucocyte-depleted components or CMV-negative cellular components are administered in these patients if they are CMV negative. CMV-negative cellular components are also administered in all patients with haematological malignancy until their CMV status is known. CMV-negative cellular components are also indicated in patients undergoing solid organ transplantation when both the donor and recipient are CMV negative.

Adverse effects of transfusion of blood and blood products

It is the recommended clinical practice that every patient is monitored closely during the first 15–30 minutes of the infusion of each unit of blood. This enables the early detection of the clinical manifestations of severe acute reaction due to incompatibility or bacterial infection, when the infusion is stopped and the necessary action taken (see below).

Within the UK, the Serious Hazards of Transfusion (SHOT) group set up a voluntary confidential reporting system for serious adverse effects excluding infectious complications in 1966. The objectives of this confidential reporting system are (1) to increase awareness of the serious adverse events, (2) to determine the factors which contribute to their occurrence (risk analysis) and (3) to issue reports based on the centralized database that recommend the necessary changes in practice designed to improve the safety of blood transfusion. The practice of SHOT is equivalent to that adopted by high-risk industries (e.g. aerospace) over many years to prevent accidents due to both malfunction and human error, thus achieving the ALARP (risk is As Low As is Reasonably Possible) endpoint. The medical co-ordinator of SHOT is Dr Elizabeth Love at the Manchester Blood Centre (http://www.shotuk.org/home/). The reporting service set up by SHOT is not designed to replace the advice and investigation of any adverse event by the local blood transfusion centre. Infectious complications must be reported to the local blood transfusion centre as soon as the diagnosis is suspected. Complications that develop in patients receiving fractionated plasma products should be reported to the Committee for Safety of Medicines using the existing yellow card system.

As the mechanisms involved in transfusion-related complications are varied, these are usually classified under separate aetiological categories. In addition, some of the adverse effects are immediate (acute), whereas others are late (Box 7.5).

Non-haemolytic febrile transfusion reactions

The routine establishment of quality control in the manufacture of both intravenous fluids and disposable giving sets has virtually eliminated pyrogenic reactions. Pyrexia following blood transfusion is nowadays the result of alloimmunization to leucocyte and platelet antigens in patients who have been immunized by previous blood transfusion or pregnancy (antibodies to class I HLA antibodies). The symptoms are rigors followed by fever, usually within 30–60 minutes after the start of the transfusion. The management includes cessation or slowing of the transfusion and the administration of an antipyretic such as paracetamol.

BOX 7.5 Complications of transfusion of blood and blood products

- Acute
  - Allergic
  - Anaphylaxis
  - Haemolytic
  - Metabolic
  - Transfusion-related acute lung injury (ARDS)
  - Circulatory overload
  - Non-haemolytic febrile transfusion reactions
  - Haemostatic: dilution of clotting factors and thrombocytopenia
  - Septic shock (bacterially infected units)
- Late
  - Delayed haemolytic transfusion reactions
  - Sensitization/alloimmunization
  - Haemolytic disease of the newborn
  - Immune suppression
  - Graft-versus-host disease
  - Transfusion iron overload (haemosiderosis)
- Disease transmission
  - Bacterial (brucellosis, syphilis, Chagas disease)
  - Helminthic (filariasis)
  - Protozoal (babesiosis, kala azar, malaria, trypanosomiasis, toxoplasmosis)
  - Rickettsial (relapsing fever, Rocky Mountain spotted fever)
  - Viral (human parvovirus B19, CMV, Epstein–Barr virus, HIV-1 and -2, human T-cell leukaemia virus (HTLV) I and -II, hepatitis A, B, G and yellow fever)

Patients who require repeated transfusions and who have experienced non-haemolytic febrile transfusion reactions are pretreated with paracetamol 1.0 g orally before the start of the transfusion, with a second dose 4 hours later. The patient is kept warm and the transfusion is administered slowly (4 hours for red cells and 2 hours for platelets). Leucocyte-depleted cell components (buffy coat-depleted red cells or filtered red cells; apheresis platelets) are given if the above measures fail.

Allergic reactions and anaphylaxis

The other non-haemolytic reactions include *allergic reactions* and *severe anaphylaxis*. The symptoms of allergic reactions are urticarial rash and itch within minutes of the start of the transfusion. Treatment is with antihistamines (chlorpheniramine 10 mg by slow intravenous or intramuscular injection) and reduction of the transfusion rate with observation of the patient. If symptoms do not progress over the next 30 minutes the transfusion is continued. Patients who had previously experienced allergic reactions to transfusion are pretreated with chlorpheniramine 8.0 mg orally 30 minutes before the transfusion.

Severe anaphylaxis is a rare (1:20 000 transfusions) but potentially fatal reaction. Occasionally, it is caused by antibodies to IgA in patients who have extremely low levels of this immunoglobulin in their plasma. Whatever the cause, anaphylaxis results in the release of vasoactive peptides and activation of complement with the development of profound hypotension, laryngeal oedema and/or bronchospasm and cutaneous flushing. Anaphylaxis is treated with immediate termination of the transfusion, intravenous crystalloids, maintenance of airway, oxygen, epinephrine, intravenous antihistamines and salbutamol. Epinephrine (0.5–1.0 mg) is given immediately by the intramuscular route and the dose repeated, if necessary, every 10 minutes, depending on the improvement in the blood pressure and pulse. Chlorpheniramine 10–20 mg is administered by slow intravenous injection after the epinephrine treatment and is continued for 24–48 hours. Salbutamol is administered by nebulizer. Because of its delayed action, hydrocortisone is of secondary value in this severe complication but its (controversial) use may prevent further deterioration. Severe anaphylaxis can only be predicted in patients with low serum IgA. Transfusion should be avoided, if at all possible, in these patients. If deemed essential they can only receive special products.

Acute haemolytic reactions

These are usually the result of ABO incompatibility resulting from human error at the bedside (blood given to the wrong patient) or in the laboratory (faulty cross-matching). The transfused cells react with the patient's own anti-A or anti-B antibodies or other alloantibodies to red cell antigens.

Incompatible blood transfusion is a serious complication and carries an average mortality of 3%, or higher if more than 200 mL of incompatible blood is administered. The reaction is usually most severe if group A red cells are administered to a group O patient. The syndrome is caused by the release of the polypeptide products of complement (C3a, C4a, C5a) in the plasma. These cause contraction of smooth muscle and degranulation of mast cells with release of vasoactive peptides (bradykinin and serotonin). Procoagulant substances are released from the stroma of the lysed red cells. These, together with antigen–antibody complexes, initiate DIC. The clinical features in the conscious patient include pain at the infusion site and along the vein, facial burning, chest and back pain, fever, rigors and vomiting. The patient becomes restless, breathless, flushed and hypotensive, and develops oozing from vascular access sites and wounds (DIC). The only manifestations of incompatible blood transfusion in unconscious or anaesthetized patients are sudden hypotension and bleeding due to DIC. The extensive intravascular haemolysis results in haemoglobinaemia and haemoglobinuria. Oliguria rapidly supervenes and progresses to ARF. The differential diagnosis is between incompatibility and the infusion of bacterially contaminated blood.

The management entails immediate recognition *with cessation of the transfusion and replacement of the giving set.* Adequate hydration is maintained by intravenous infusion of crystalloids (isotonic saline or Hartmann's solution). An attempt is made to force diuresis with intravenous large-dose furosemide (150 mg). If this fails, a 20% solution of mannitol (100 mL) is administered. If diuresis is obtained, a high urine output is maintained (100 mL/h) by large-volume crystalloid infusions. Often, however, these patients progress to ARF, necessitating haemodialysis. The other problems that require immediate support are bleeding from DIC (blood component therapy guided by clinical state and coagulation screen) and hyperkalaemia. Intravenous glucose–insulin (50 mL 50% glucose + 10 units of insulin) is administered if the serum K^+ rises above 6.0 mmol/L. This is followed by an intravenous infusion of 10% glucose containing 10 units of insulin over a period of 4 hours. *After the initial resuscitation is completed, the investigation of such an incident is essential and is outlined in Box 7.6.*

Acute haemolytic reactions giving a similar picture may arise from acute haemolysis caused by preformed antibodies in the patient's blood (anti-Rh (D), Rh (E), Rh (C) and Kell, etc.) as a result of alloimmunization to minor blood group antigens in the donated unit. These may be encountered in patients requiring repeated blood transfusions. Those caused by anti-Rh (D) (rare nowadays since patients receive Rh (D)-compatible cells) are not usually severe as they do not activate complement.

Delayed haemolytic transfusion reactions (DHTRs) are rare but can occur in patients whose level of antibodies to the blood group antigen is so low that it escapes detection by the pretransfusion screen. Following transfusion, the secondary immune response raises the antibody titre to a level that results in the delayed destruction of the transfused cells. Thus, the manifestations, which include fever, falling Hb, jaundice and haemoglobinuria, appear some 5–10 days after the transfusion. DHTRs are seldom fatal.

Transfusion-related acute lung injury

TRALI is one cause of ARDS and was previously thought to result from pulmonary microvascular occlusion by microaggregates of platelets, leucocytes and fibrin (50–200 μm), which are known to be present in stored blood. For this reason, microaggregate filters were recommended for transfusions in excess of 5 units of blood. A more definite cause is donor blood containing antibodies to the patient's leucocytes (nearly always donations from multiparous women). Following transfusion, the patient develops fever, increasing breathlessness, non-productive cough and hypoxaemia. The chest radiograph shows the typical features of ARDS with perihilar infiltrates leading to a whiteout in severe cases. The management is that of ARDS.

Metabolic and haemostatic complications

These complications are confined to patients who, because of severe haemorrhage, receive a massive blood transfusion of stored blood. *Massive blood transfusion is defined as that which is equivalent to or exceeds the patient's own blood volume within 12 hours.* Aside from being cold (4°C), stored blood has an acid pH and a reduced 2,3-DPG, and contains citrate anticoagulant, elevated plasma potassium and ammonia. The metabolic consequences therefore include hypothermia, acidosis and an increased affinity of oxyhaemoglobin for oxygen that is not readily released to the tissue, thereby contributing to defective tissue oxygenation. Hypothermia may lead to cardiac arrhythmias including ventricular fibrillation and asystole. For this reason, blood warming is necessary if the transfusion rate exceeds 50 mL/min. Unfortunately, the heating coils increase the resistance of the giving circuit, but nonetheless their use is essential in these patients. Citrate intoxication is due to the chelation of ionized calcium, which may result in prolongation of the QT interval, but this does not usually materially affect cardiac function and the ionized calcium levels rapidly return to normal after the transfusion as the excess citrate is metabolized and excreted. Thus, the use of supplemental calcium is not justified, particularly as it may itself give rise to arrhythmias. Hyperkalaemia is seldom a problem as the excess plasma K^+ enters the RBCs with warming to body temperature. It is, however, a consideration in patients with acidosis and renal failure when calcium is administered as the physiological antidote.

BOX 7.6 Investigation of an acute haemolytic transfusion episode

1 Report incident to the blood transfusion department
2 Establish that the unit has been issued to the patient receiving it
3 Obtain fresh samples of patient's blood (clotted and in EDTA) for re-cross-matching and serological testing
4 Send sample of the blood unit for culture
5 Return unit to the blood transfusion department
6 Obtain further blood samples from the patient for clinical chemistry (electrolytes, urea, free Hb) and coagulation screen
7 Discuss any further transfusion requirements with the blood transfusion officer

Stored blood is deficient in platelets and labile clotting factors (V and VIII). For this reason, massive transfusion of stored blood induces a dilution of the labile clotting factors in addition to a moderate thrombocytopenia. The deficiency of the labile clotting factors can be circumvented by the administration of 2 units of FFP for every 8 units of blood.

Platelet-specific alloantibodies may cause *post-transfusion purpura*, which is most commonly encountered in females and may prove fatal. The syndrome becomes manifest some 5–9 days after the transfusion with severe bleeding associated with extreme thrombocytopenia. Initially treatment is with high-dose corticosteroids and high-dose intravenous immunoglobulin. If platelets are needed they have to be compatible with the patient's alloantibodies.

Circulatory overload

This is encountered when blood is administered too rapidly or in large volumes. It is most commonly encountered during the transfusion of anaemic patients, particularly those with severe and chronic anaemia. These patients must be transfused very slowly and only with packed cells (with or without concomitant diuretic therapy). In addition, transfusion is restricted to 1 unit in any 12 hour period. In some patients, an exchange transfusion has to be carried out to avoid severe congestive failure.

Pulmonary oedema consequent on left ventricular failure is common after massive blood transfusion and must be differentiated from ARDS.

Immune suppression

There is no doubt concerning the immunosuppressive effect of blood transfusion which, indeed, was employed specifically for this purpose in patients before renal transplantation to improve graft survival prior to the introduction of cyclosporin, the effective immunosuppression of which has made prerenal transplant transfusion unnecessary. In the context of general surgery, perioperative blood transfusion, by virtue of this immunosuppressive effect, which is additive to that inherent to the operative trauma, has undoubted undesirable consequences. Aside from the risk of circulatory overload, perioperative blood transfusion enhances the risk of infective complications (proven), and may increase recurrence rate and reduce disease-free survival in patients after extirpative surgery for cancer. Although the latter is unproven, there is circumstantial evidence for this adverse effect, especially in patients undergoing resections for colorectal cancer.

Transfusion haemosiderosis

Every unit of blood contains 250 mg of iron that is retained by the body. Iron overload of the monocyte–macrophage system is caused by repeated red cell transfusions over many years and becomes significant after 100 units have been administered, when the liver, pancreas, myocardium and the endocrine glands become damaged. It is especially a problem in childhood anaemias (e.g. thalassaemia) and in patients with chronic refractory anaemia. The iron overload is reduced in these patients by iron chelation therapy with desferrioxamine.

Graft-versus-host disease

This is a rare but usually fatal complication that occurs mainly in immunodeficient patients, e.g. recipients of allogeneic marrow transplants and fetuses receiving intrauterine transfusions. However, graft-versus-host disease has also been documented in immunocompetent patients after transfusion of blood from a relative. The disease is caused by the T-lymphocytes and starts some 4–30 days after the transfusion. The patient develops a high fever and a diffuse erythematous skin rash progressing to erythroderma and desquamation, GI symptoms, severe hepatic dysfunction and pancytopenia. Graft-versus-host disease is prevented by administering gamma-irradiated cellular components to immune-deficient patients. Similarly, blood donated from relatives should be gamma irradiated.

Transmission of infectious disease

A wide spectrum of infectious diseases can be transmitted by the transfusion of blood and blood products, although, with the modern practice of screening of blood donors and heat treatment of blood protein products, the risk is extremely small. The most commonly transmitted viral disease is hepatitis C (<1 per 30 000 components transfused). Post-transfusion hepatitis B is now extremely rare in the UK. Hepatitis A can very occasionally be transmitted by blood products. Hepatitis G has only been recently identified. Currently, it can only be identified by gene amplification technology and there is no screening test. It is not known whether hepatitis G can cause serious disease and whether the existing plasma fractionation and heat treatment inactivate it, although this is thought likely.

Human parvovirus B19 may not be inactivated by current plasma fractionation and heat treatment. It causes depressed erythropoiesis in some patients. Human T-cell leukaemia virus (HTLV)-related disease following transfusion is extremely rare in the UK and, for this reason, donors are not screened for HTLV-I/II infection. However, the prevalence of HTLV-I is high in some countries, notably Japan and the Caribbean. HTLV-I causes neurological disease and a rare form of adult T-cell leukaemia, usually for many years after the transfusion.

CMV is a problem as 50% of UK donors have antibodies to CMV but fortunately only a fraction of antibody-positive donations transmits the virus. Post-transfusion CMV is important in premature infants born to CMV antibody-negative mothers and in CMV-negative recipients of bone marrow allografts from CMV-seronegative donors. These patients should receive CMV-negative products or leucocyte-depleted blood components. The HIV problem has largely been resolved by donor selection and testing.

Transfusion of bacteriologically contaminated or infected blood

This disastrous complication is fortunately rare in the UK (2 per million units transfused). The pathogens are usually cold-growing strains of *Pseudomonas fluorescens* or *Yersinia enterocolitica*. Skin organisms such as staphylococci can proliferate in platelet concentrates stored at 20–22°C. The clinical picture is similar to that of ABO-incompatible blood transfusion reaction. Despite aggressive supportive therapy, the majority of patients do not survive this complication.

Solutions for intravenous infusion

These include colloid solutions (ALBA 4.5% and synthetic colloids), isotonic crystalloid solutions, hypertonic saline solutions and red cell substitutes. There has been debate regarding the comparative efficacy of crystalloids versus colloids for volume replacement in the resuscitation of critically ill patients. A meta-analysis of published randomized trials (Schierhout and Roberts, 1998) indicated that resuscitation with colloids is associated with an increased absolute risk of mortality of 4%, i.e. four extra deaths for each 100 patients resuscitated. *This systemic review does not support the continued use of colloids.*

Synthetic colloid solutions: plasma expanders

Synthetic colloids are preferred to ALBA during resuscitation of hypovolaemic patients (except in burn patients), largely on the grounds of costs. Thus, for example, 1 unit of ALBA 4.5% costs £35, whereas the equivalent 500 mL solution of gelatin retails at £3–4. Synthetic colloidal solutions are used, as an alternative to isotonic crystalloids, in the resuscitation of hypovolaemic patients before blood is available or when it is unnecessary. Three types of preparation are available: dextrans, gelatins and hetastarch (HES) (Table 7.7).

The clinical effects of a synthetic colloid depend on its colloid osmotic pressure, its plasma half-life and its capillary permeability. The colloid osmotic pressure is expressed by the molecular weight and concentration of the colloid, but, as these are non-uniform solutions (particles are not of uniform weight), their molecular weight is better expressed by the 'number average' (M_n) rather than 'weight average' (M_w). M_n represents the mean osmotically active particle weight and is compared with the molecular weight of albumin (70 000).

Dextrans

These are solutions of inert glucose polymers produced from hydrolysed starch, which are available in various molecular sizes (dextran 40, dextran 70, etc.). As the renal threshold for these inert polysaccharides is of the order of 50 000, dextran 40 molecules are filtered through the glomeruli, whence they may block the renal tubules and cause ATN, especially in patients with reduced renal blood flow. For this reason, dextran 40 is rarely used nowadays. Dextran 70 is free from this side effect and can be used for volume replacement (up to 1.5 L/24 h in an adult). It is available either in saline (0.9%) or in 5% dextrose (when added Na$^+$ is undesirable). The colloid osmotic pressure of dextran 70 is 268 mmH$_2$O (lower than those of other plasma substitutes), but, because of its half-life (12 hours), its effect is relatively prolonged.

Dextran 70 infusion affects the coagulation system by inhibiting platelet aggregation and rendering fibrin more susceptible to fibrinolysis. For these reasons, dextran 70 is used in some centres in the prophylaxis of postoperative deep vein thrombosis. Large infusions (more than 20 mL/kg/day) can reduce factor VIII levels and induce capillary oozing. This limit of 1.5 litres of dextran 70 must not be exceeded during volume replacement. As dextran 70 interferes with blood cross-matching by inducing rouleux formation *in vitro,* a blood sample should be obtained before the infusion is commenced.

Adverse effects of dextran infusion are well documented. These are usually mild anaphylactoid reactions (1:2000 infusions) but serious reactions can occur, although they are rare (1:6000 infusions).

Hetastarch (Hespan)

This is not frequently used in the UK but is widely employed in the USA, where gelatin preparations are not available. It consists of a 6% solution of HES (chemically modified maize starch) in 0.9% saline. Its average particle size ($M_n = 70\,000$) and colloid osmotic pressure are equivalent to those of 5% albumin. Although the smaller molecules are removed by glomerular filtration (40% within 24 hours), the larger polymers undergo slow hydrolysis by plasma α-amylase, and some are then filtered once the size is reduced below the renal threshold. Approximately 30% of the infused dose ultimately leaves the vascular compartment and is taken up by the monocyte–macrophage system, predominantly in the liver and spleen. There are some concerns over the potential for impaired function (blockade) of this system following large infusions of hetastarch. In any event, the final elimination from the body is a slow process, and, at 4 months, the plasma concentration averages 1% of the original value. The half-life of hetastarch in the plasma is considerably longer than any of the other plasma substitutes and averages 17 days. In animals, infusions of HES exceeding 20 mg/L (equivalent to 1500 mL in a 70 kg patient) impair platelet function and coagulation. Hetastarch administration may be followed by a rise in the serum amylase but there is no evidence of any pancreatic damage. In a recent report, one serious and 13 mild reactions were encountered in 16 000 infusions.

Table 7.7 Synthetic colloid solutions: plasma expanders

Product	Nature	M_n	COR (mmH$_2$O)	Half-life
Dextran 70	Hydrolysed starch	70 000	268	12 hours
Hetastarch, Hespan, HES	Hydroxyethyl maize starch	70 000	350	17 days
Gelofusine	Succinylated gelatin	22 600	465	4 hours
Haemaccel	Polygeline	24 500	350–390	5 hours

COR, colloid osmotic pressure; M_n, number average.

Gelatin solutions

The modern preparations consist of solutions of heat-degraded cattle bone gelatin: succinylated gelatin (Gelofusine) and polygeline (Haemaccel). Although both are made up in balance crystalloid solution, the calcium concentration is different. Haemaccel contains over 10 times more calcium (6.25 mmol/L) and potassium (5.1 mmol/L) than Gelofusine. The high calcium content of Haemaccel can lead to clotting in the warming coils when this colloid is mixed with citrated blood or FFP. Whereas the colloid osmotic pressure of Haemaccel (350–390 mmH$_2$O) is similar to that of plasma, that of Gelofusine is greater (465 mmH$_2$O), i.e. it tends to draw fluid into the intravascular compartment from the interstitial space. As their particle weight is small (Haemaccel $M_n = 24\,500$; Gelofusine $M_n = 22\,600$), they are rapidly filtered by the renal glomeruli and cause an osmotic diuresis which is beneficial. This too accounts for their short half-life: 5 hours for Haemaccel and 4 hours for Gelofusine. As neither preparation affects either the bleeding time or coagulation time (except by dilution), they are perfectly safe for large-volume colloid replacement. There is some evidence that some of the gelatin complexes may leave the vascular compartment and become sequestered in the interstitial space, particularly the lung. These extravasated complexes may contribute to the increased lung water, which may retard the return of pulmonary gas exchange to normal.

The original version of Haemaccel led to a number of histamine-induced adverse reactions when the solution was infused rapidly, but, since 1981, the manufacturing process has been altered to reduce the amount of cross-linking agent responsible for the plasma histamine released. The incidence of minor reactions with the modified Haemaccel is nine per 1334 infusions. The reported serious anaphylactoid reactions with Gelofusine infusion vary from 1:6000 to 1:12 000. Within the UK, the concern with gelatin solutions is their bovine origin, possible contamination with the prion protein responsible for bovine spongiform encephalopathy (BSE) and the possible development of new-variant Creutzfeldt–Jakob disease but, to date, no health warnings have been issued. Presumably, the heat-degradation process destroys the prion protein. It seems likely that the original gelatin will be obtained from BSE-free herds in the future.

Crystalloid solutions

These are isotonic balance salt or dextrose solutions (Table 7.8). The two most commonly used for volume replacement are isotonic saline (0.9%) and Ringer's lactate (Hartmann's solution), which approximates most closely to the electrolyte composition of plasma. This and the less acidic pH of Hartmann's solution (6.5 versus 6.1) account for its preferment over isotonic saline by some surgeons, while others consider the lactate content of Hartmann's solution to be a disadvantage since shock is accompanied by elevated blood lactate, and, for this reason, prefer isotonic saline. Amid this minor controversy, all are agreed that 5% dextrose has no place in resuscitation. It is equivalent to the safe intravenous administration of water and the little dextrose it contains is not metabolized. As distinct from colloid following intravenous infusion, crystalloid solutions are rapidly distributed between the intravascular and interstitial spaces roughly in the proportion of 3:1. This is not a disadvantage, as in shock both compartments are contracted. However, the amount required to restore the CBV is significantly larger (by at least a factor of 3) then when colloid solutions are used. Thus, the replacement of 1.0 litre of circulating blood requires the infusion of 3.0 litres of Hartmann's or isotonic saline, as opposed to 1.0 litre of colloid. The extent of expansion of the body fluid compartments limits the magnitude of losses that can be replaced solely by crystalloid solutions and crystalloid solutions predispose to oedema and the compartment syndrome.

Combined crystalloid–colloid blood replacement

The current widely practised compromise for blood volume replacement entails the infusion of a mixture of crystalloids and synthetic colloids in a ratio of 2:1. Stored blood or red cell concentrates are added when the blood loss exceeds 1.5 litres or if the haematocrit falls below 25%.

Hypertonic saline and hypertonic saline solutions

Intravenous infusions of hypertonic crystalloid solutions (500 mmol/L) have been reported to have beneficial effects in patients with refractory shock, in whom they may increase the survival rate. Several formulations have been proposed. The Sackford solution contains the following in mmol/L: Na^+ 250, Cl^- 100, lactate 63, K^+ 4 and Ca^{2+} 9. These hypertonic solutions increase CO and possibly improve cerebral oxygen delivery in severely shocked patients. The exact mechanism of action is unknown but is possibly vagally mediated by the arrival of the hypertonic bolus in the pulmonary vasculature. Some reports indicate that hypertonic solutions reduce the peripheral resistance and are accompanied by a lesser weight gain than ordinary isotonic crystalloid solutions. The duration of the effect on the circulation of hypertonic solutions is short lived and probably does not exceed 1 hour. Although there is good experimental evidence from animal studies of the benefit of hypertonic saline/hyperoncotic solutions in hypovolaemia due to blood loss from rapid expansion of blood volume, improvement in cardiac index and stimulation of the immune system with improvement in survival, the published studies on use of these solutions in humans remain limited, lack appropriate controls and there have been no randomized studies. A major publication (Oliveira *et al*. 2002) has reviewed the published experimental evidence on the use of these hypertonic solutions in sepsis and septic shock and concluded that they appear to have the potential to benefit patients with sepsis or septic shock, and possibly in the prophylaxis of septic shock and the systemic inflammatory response syndrome.

Table 7.8 Crystalloid solutions for intravenous use (mmol/L)

Solution	Na^+	K^+	Cl^-	HCO_3^-	Ca^{2+}
Isotonic (N) saline	154	–	154	–	–
Dextrose 5% N saline	154	–	154	–	–
Dextrose 5% N/2 saline	77	–	77	–	–
Hartmann's solution (Ringer lactate)	131	4	110	(28)*	3
Ringer's solution	147	4	155	–	4
Dextrose 5%	–	–	–	–	–

*As lactate.

Several clinical studies in trauma patients have indicated potential benefits, although the reported data are not conclusive. In one study, stable patients with septic shock were administered 2–4 mL/kg hypertonic saline in HES. This resulted in minor improvement in oxygenation through an increased CO and oxygen delivery.

In another study, administration of 250 mL hypertonic saline in dextran significantly improved the cardiac index and pulmonary artery occlusion pressure compared with administration of an equivalent volume of normal saline. Hypertonic solutions are known to stimulate the immune system and exhibit an anti-inflammatory effect, which may prevent or minimize the severity of multiple organ failure. The probable advantages of using hypertonic solutions in patients with sepsis include: (1) rapid and sustained expansion of plasma volume with improved cardiovascular function, (2) restoration of microcirculatory perfusion, (3) improved immune function and (4) reduced bacterial translocation from the gut. However, prospective clinical studies are needed to confirm the safety and efficacy of hypertonic solutions in clinical practice, and especially in the treatment and prophylaxis of septic shock.

Red cell and platelet substitutes

In principle artificial blood products are needed to address the chronic deficiency relative to the clinical needs of blood and blood products obtained from voluntary human donations. In addition, these substitutes have certain benefits.

Artificial blood products have the following benefits:

- readily available and have a long shelf-life
- they can be filtered and pasteurized to eliminate microbial contamination and disease transmission
- they do not require blood typing, so they can be infused immediately
- they do not cause immunosuppression in the recipient.

Blood substitutes

Red cell substitutes, despite considerable research over many years, are still in the development phase and their use in clinical practice, largely limited to clinical trials, has become possible only recently. Blood substitutes are either Hb based or synthetic perfluorocarbons (PFCs).

Haemoglobin substitutes

The modified Hb solutions consist of polymerized Hb complexed with pyridoxine-5-phosphate and dialysed to remove vasoconstrictor substances present in the red cells. When Hb is extracted from the RBCs, it rapidly loses its 2,3-DPG, which is necessary for its low affinity for oxygen (essential for the release of oxygen to the tissues). This problem is overcome by complexing the Hb with pyridoxine-5-phosphate, which restores the normal oxygen affinity. Polymerization is necessary to increase the molecular size as non-polymerized Hb rapidly leaves the vascular compartment and is filtered through the glomeruli into the urine. However, the previously held view that Hb is nephrotoxic has been shown to be incorrect. Renal damage is caused by the red cell stroma and not the released Hb. The polymerization process is also viricidal and results in a fourfold increase in the half-life of the Hb (20 hours). Modified Hb solutions contain up to 150 g/L of Hb.

An alternative approach is to encapsulate the Hb/pyridoxine-5-phosphate in lipid membrane microspheres of equivalent size to the RBCs. The half-life of solutions made of Hb-containing microspheres is similar to that of polymerized Hb solutions.

The Hb substitutes are made from human, animal or recombinant Hb. There are justifiable concerns regarding animal Hb substitutes as this is obtained from cows' blood and there is a risk of transmission of prions that cause BSE. Recombinant Hb is produced following insertion of the human gene into bacteria followed by isolation and purification from the bacterial culture. The problem with all Hb substitutes is that the infused Hb rapidly breaks down from its tetramer configuration to dimers and monomers. Thus, stabilization of the Hb tetramer structure is necessary and is achieved by cross-linking with sugars and certain polymers or by pyridoxylation. The latter creates a substitute with near-normal oxygen binding. Another problem is that free Hb binds with nitric acid, and this complex then causes vasoconstriction and hypertension.

Hb can be polymerized with surface amino acid groups by glutaraldehyde. Polymerized Hb is the only substitute that does not cause significant vasoconstriction after infusion and has been used successfully in phase III clinical trials. Hemolink is another substitute created from human Hb polymerized with oraffinose. PolyHeme, also produced from human Hb, has been used in phase III trials in trauma patients, who received up to 10 units of PolyHeme with minimal side effects. In contrast, Hemopure is made from bovine Hb. A multicentre trial in patients undergoing infrarenal aortic reconstruction reported that 27% of patients receiving Hemopure were able to avoid transfusion of allogeneic blood.

Perfluorocarbon substitutes

In contrast to the chemical reaction involved in oxygen transport by Hb, emulsions of PFCs carry oxygen in physical solution at high partial oxygen pressures, although they readily release the oxygen when the PO_2 is low, i.e. at tissue level. The need for high partial oxygen pressure required for adequate oxygen uptake from the lung limits their clinical usefulness, although some of the newer formulations have overcome this problem to some extent. PFCs have a short intravascular half-life and, following extravasation, are taken up by the cells of the monocyte–macrophage system. One advantage of these solutions is their low viscosity, which offers advantages in some clinical situations, e.g. limb salvage and patients with myocardial infarction. Recent solutions consist of combinations of a mixture of three PFCs with HES. The 20% solution requires inspired oxygen of 100%, but the 35% solution can be used with 60% inspired oxygen, rendering it more acceptable for general clinical use.

PFC-based solutions have the capacity to dissolve up to 50 times more oxygen than plasma. The problem is that PFC solutions do not mix well with blood and thus have to be emulsified with lipids. Additionally, patients have to be on 100% oxygen during infusion ($PaO_2 \geq 350$ mmHg). The PFC-based solutions include Oxygent, which is currently in phase III clinical trials in cardiac and general surgical patients. It can be used for augmented controlled normovolaemic haemodilution before major surgery as a means of avoiding allogeneic blood transfusion. PFC-based substitutes are acceptable to Jehovah's Witnesses.

Adverse reactions to blood substitutes

The following indicate their limitations and the need for improved substitutes:

- Hb-based substitutes:
  - elevations in blood pressure
  - gastrointestinal dysmotility
  - mild, transient increase in pancreatic enzymes
  - jaundice
  - interfere with laboratory testing
  - short duration of action – 24 hours
  - expensive.
- PFC substitutes:
  - mild thrombocytopenia (10–15%)
  - flu-like syndrome
  - oxygen toxicity with prolonged use as the patients have to be on 100% inhaled oxygen
  - short action – 24 hours
  - expensive.

Platelet substitutes

The two problems necessitating the development of platelet substitutes are (1) short shelf-life, typically 5 days, and (2) need for storage at room temperature with the risk of bacterial overgrowth. The various platelet substitutes under development include:

- infusible platelet membranes (IPMs)
- thrombospheres
- lyophilized human platelets.

Infusible platelet membranes

These are produced from outdated human platelets which are fragmented, virally inactivated and lyophilized, enabling storage for up to 2 years. They are currently in phase II clinical trials. IPM infusions are indicated in patients who have become refractory to platelet transfusions because of the formation of antiplatelet antibodies. IPMs appear to be clinically effective in stopping bleeding and are safe.

Thrombospheres

Thrombospheres (Hemosphere) are composed of cross-linked human albumin bound to human fibrinogen. They are still in the experimental phase of development but studies indicate that they are effective in enhancing platelet aggregation. A similar product (Synthocytes) is undergoing clinical trials in Europe.

Lyophilized human platelets

Lyophilized platelets are produced by fixing human platelets in paraformaldehyde prior to freeze-drying in an albumin solution. The adhesive properties of the platelets appear to be maintained. This product is not as yet in clinical use.

Oedema

Oedema is defined as the excessive accumulation of tissue fluid, mainly in the interstitial but also in the intracellular space. Under physiological conditions, the Starling forces (capillary pressure and plasma oncotic pressure) ensure the orderly bidirectional movement of fluid between the intravascular compartment [more specifically the effective circulatory volume (ECV)] and the interstitial space. Thus, during the course of 1 day, some 20 litres of fluid leave the capillaries at the arteriolar end and 18 litres are absorbed back across the capillary membrane at the venous zone of the capillary network, leaving an excess of 2 litres, which is returned to the circulation by the lymphatics. Oedema represents a breakdown of this balanced fluid transport with a net accumulation of tissue fluid.

Oedema may be *localized* (e.g. inflammatory, neurogenic, local hypersensitivity reaction), when only an imbalance of the Starling forces and other locoregional factors (e.g. increased vascular permeability) are operative, or *generalized*, when the role of the kidney is of paramount importance, irrespective of the aetiology (cardiac, hepatic or renal disease).

Mechanisms of generalized oedema formation

Chronic disease

The basic mechanisms that are common to all disorders accompanied by excessive water and salt retention causing oedema with or without ascites or hydrothorax are outlined schematically in Figure 7.8. In addition, other factors that are disease specific may be involved, e.g. hypoalbuminaemia in chronic liver and renal disease and relative excess of aldosterone (secondary hyperaldosteronism) in cirrhosis, but these are less important in the aetiology of interstitial tissue fluid overload.

The common causes of oedema are:

- pregnancy
- renal disease
- cardiac failure
- chronic pulmonary disease
- thyroid disease
- chronic liver disease
- diabetes
- arthritis
- malnutrition
- drugs, e.g. corticosteroids, calcium channel blockers for hypertension
- contraceptive pill.

In heart failure, chronic hepatic and renal disease, the raised capillary pressure is due to an elevated venous pressure that is transmitted to the capillary bed. This reduces the amount of fluid reabsorbed from the interstitial space back into the intravascular compartment. The net result is a contraction of the plasma volume which, in itself, would tend to limit the pathological process. However, the low-pressure volume receptors in the atrium and pulmonary vessels detect a reduction in the ECV and trigger the excess reabsorption of water and NaCl by the kidneys, so that the plasma volume is restored. The net outflow of fluid to the interstitial space is continued until a steady state is reached. This occurs when the pressure in the interstitial space is raised sufficiently to counter the elevated capillary pressure. In essence, this means that tissue perfusion is impaired.

In some disorders, the oncotic pressure is reduced. This may be due to the reduction in the serum albumin from increased permeability of the glomerular capillaries in certain forms

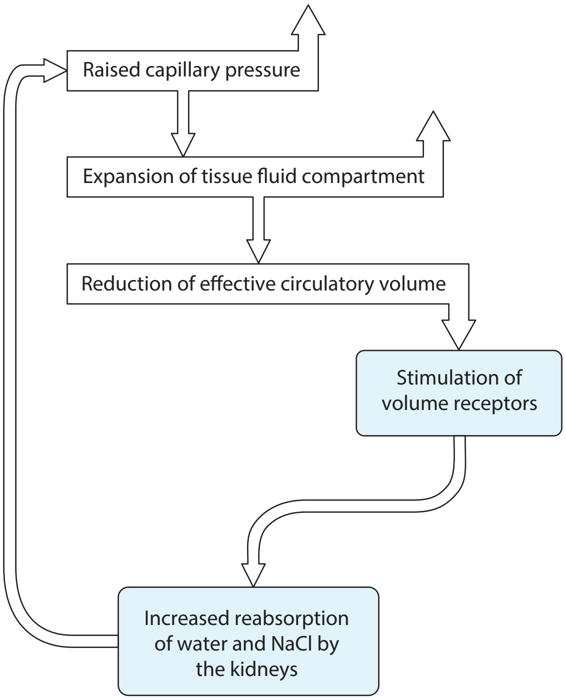

Raised capillary pressure
Expansion of tissue fluid compartment
Reduction of effective circulatory volume
Stimulation of volume receptors
Increased reabsorption of water and NaCl by the kidneys

Figure 7.8 Basic mechanisms involved in generalized oedema.

of renal disease (e.g. nephrotic syndrome) or to defective synthesis (e.g. chronic liver disease) or protein malnutrition (e.g. kwashiorkor). The reduced oncotic pressure augments the effect of the raised capillary pressure in the net accumulation of tissue fluid in liver and kidney disease.

Clinically, oedema is often dependent, and, for this reason, is commonly found around the ankles in ambulant patients or over the sacral area when the patient is bed-bound. The cause of this distribution can be directly ascribed to the greater venous (and hence capillary) pressures in these regions and to the effects of gravity. Oedema is also prominent in patients with renal disease in the periorbital regions, where the tissues are lax. Certain individuals, usually female, suffer from a deficiency of renal handling of sodium and this leads to cyclical oedema manifesting in puffiness of the hands, feet and face.

Critically ill patients

Generalized oedema is often encountered in critically ill patients on ventilatory and inotropic support (major trauma and advanced sepsis). This is usually multifactorial, e.g. excess crystalloid infusion with pathological expansion of the interstitial space, third space losses, low CO, increased capillary permeability and low plasma oncotic pressure (dilution). The concept of the 'third space' arose from the observation that surgical trauma and inflammatory disease resulted in increased capillary permeability and sequestration of extracellular fluid in the traumatized or inflamed tissue (e.g. oedematous bowel in intestinal obstruction or peritonitis). The generalized oedema of critically ill patients responds to intravenous fluid restriction (haemodynamic monitoring permitting) if the patient's condition improves. Diuretics are seldom indicated as most of these patients are in renal failure. Haemodialysis with hyperosmolar dialysate may be necessary.

Mechanisms of localized oedema formation

Here, the important pathological variables are vasodilatation and increased permeability of the capillary endothelium. The vasodilatation of the precapillary sphincters results in elevation of the capillary pressure with excess transudation over reabsorption across the capillary membrane.

Inflammatory oedema

The increased permeability of the capillary endothelium, which is caused by the cytokine cascade, components of complement following activation by the classic or alternative pathway, kinins, prostaglandins, proteases (e.g. granulocyte elastase), etc., results in the escape of both fluid and albumin in the interstitial space. The extravasated albumin, by raising the oncotic pressure of the tissue fluid, further accentuates the net accumulation of fluid in this compartment. These mechanisms operate irrespective of the aetiology of the condition: bacterial inflammation, inflammatory oedema of healing wounds, oedema due to anaphylaxis, neurogenic oedema, etc.

Apart from its local and systemic effects, inflammatory oedema is important from a technical aspect in relation to tissue approximation (closure of wounds and intestinal anastomosis). Oedematous tissue sustains a marked reduction in its tensile strength and, for this reason, when the edges are approximated by sutures, there is a tendency for the sutures to cut through with breakdown of the wound or dehiscence of the anastomosis.

Limb oedema

In patients with oedema affecting one limb, the usual cause can be found in some problem within the deep venous system. Often, this is the result of previous deep venous thrombosis, which leads to the postphlebitic syndrome (marked lower limb oedema, eczema, pigmentation and ulceration). In a small proportion of patients, compression of the left common iliac vein by the right common iliac artery may be responsible for left-sided leg oedema.

Less commonly, limb swelling may arise from lymphoedema owing to blockage or insufficiency of the lymphatic trunks draining the limb. The most common variety is secondary lymphoedema from blockage or disruption due to trauma, surgical excision of lymph nodes (postmastectomy), radiotherapy for cancer, infiltration by malignant disease, parasitic infestations, e.g. filariasis, etc. In primary lymphoedema, the mechanism is insufficiency rather than obstruction (hypoplastic or incompetent lymphatics).

Whatever the cause of lymphoedema, the accumulated fluid has a high protein content, which induces a fibrotic reaction in the subcutaneous region and is particularly prone to infection, especially by streptococcal organisms. These inflammatory episodes aggravate the subdermal fibrosis so that the skin becomes thick, dry, scaly and rough (elephantiasis). Long-standing lymphoedema does not pit on pressure, although early lymphoedema does. Lymphoedema predisposes to recurrent bacterial infection of the affected limb, which aggravates the fibrosis and induration. Hence, prophylactic personal hygiene is an essential part of the management of these patients (see Chapter 11).

Other unusual causes of limb swelling include: *disuse oedema*, when the patient holds the limb completely immobile for long periods, *post-traumatic osteodystrophy* (very painful swelling such as to make the limb useless, accompanied by osteoporosis, also known as Sudeck's atrophy) and *local gigantism*, which may be due to arteriovenous malformations or Klippel–Trenaunay syndrome (combination of congenital varicose veins, abnormal deep veins and capillary naevi).

Angio-oedema

This is caused by the acquired or congenital absence of C1 esterase inhibitor. As a result of this deficiency, afflicted individuals are prone to episodes of unchecked activation of complement with marked oedema formation that often affects the upper airway, leading to life-threatening asphyxia. Angio-oedema occurs in a subset of patients with immunological urticaria. In affected patients the C1 esterase inhibitor deficiency leads to release of complement-derived kinins, which are the immediate mediators of the oedema. Acquired angio-oedema is also encountered in patients on angiotensin-converting enzyme (ACE) inhibitors owing to inhibition of kinin breakdown by ACE.

Hereditary and acquired C1 inhibitor deficiency is screened by assay of serum C4, which, if low, requires quantitative and functional assays of the C1 esterase inhibitor to confirm the diagnosis.

Treatment of C1 esterase inhibitor deficiency

Maintenance therapy is only necessary for patients with symptomatic recurring angio-oedema. Anabolic steroids are used in most patients but may cause virilizing side effects. Tranexamic acid is also used for maintenance but is contraindicated in patients with a history of thrombosis. Prophylaxis before elective surgery includes tranexamic acid started 3–4 days before the planned intervention or by increasing the dose of tranexamic acid or anabolic steroids in patients on maintenance therapy. C1 inhibitor concentrate or FFP is used for emergency treatment of serious angio-oedema attacks or as prophylaxis before emergency surgery, especially when intubation is necessary. Epinephrine is ineffective in angio-oedema caused by C1 esterase inhibitor deficiency. Acute episodes are treated by infusion of C1 esterase inhibitor derived from plasma and antihistamines. Patients may require endotracheal intubation or tracheostomy.

GUIDE TO FURTHER READING

Bellomo R, Kellum JA, Mehta R, *et al.* Acute Dialysis Quality Initiative II: the Vicenza conference. *Curr Opin Crit Care* 2002;**8**:505–508.

Blower AL, Brooks A, Fenn GC, *et al.* Emergency admissions for upper gastrointestinal disease and their relation to NSAID use. *Aliment Pharmacol Ther* 1997;**11**:283–291.

Bombardier C, Laine L, Reicin A, *et al.*, for the VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med* 2000;**343**:1520–1528.

Contreras M (ed.). *ABC of Transfusion*. London, UK: BMJ Publications, 1990.

Donnelly MT, Hawkey CJ. Review article: COX-II inhibitors – a new generation of safer NSAIDs? *Aliment Pharmacol* 1997;**11**:227–236.

Emori Y, Sakugawa M, Niiya K, *et al.* Life-threatening bleeding and acquired factor V deficiency associated with primary systemic amyloidosis. *Blood Coagul Fibrinolysis* 2002;**13**:555–559.

Faulkner G, Pritchard P, Somerville K, Langman MJS. Aspirin and bleeding peptic ulcer in the elderly. *BMJ* 1988;**297**:1311–1313.

Fry AC, Farrington K. Management of acute renal failure. *Postgrad Med J* 2006;**82**:106–116.

Galan AM, Bozzo J, Hernandez MR, *et al.* Infusible platelet membranes improve hemostasis in thrombocytopenic blood: experimental studies under flow conditions. *Transfusion* 2000;**40**:1074–1080.

Grattan C, Powell S, Humphreys F. Management and diagnostic guidelines for urticaria and angio-oedema. *Br J Dermatol* 2001;**144**:708–714.

Henry DA, O'Connell DL. Effects of fibrinolytic inhibitors on mortality from upper gastrointestinal haemorrhage. *BMJ* 1989;**298**:1142–1146.

Napier JAF. *Handbook of Transfusion Therapy*, 2nd edn. Chichester, UK: John Wiley, 1996.

Nussmeier NA, Whelton AA, Brown MT, *et al.* Complications of the COX-2 inhibitors parecoxib and valdecoxib after cardiac surgery. *N Engl J Med* 2005;**352**:1081–1091.

Oliveira RP, Velasco I, Soriano FG, Friedman G. Clinical review: hypertonic saline resuscitation in sepsis. *Crit Care* 2002;**6**:418–423.

Schierhout G, Roberts I. Fluid resuscitation with colloid or crystalloid solutions in critically ill patients: a systemic review of randomised trials. *BMJ* 1998;**316**:961–964.

Sihler KC, Napolitano LM. Massive transfusion: new insights. *Chest* 2009;**136**:1654–1667.

Stael von Holstein CCS, Eriksson SBS, Kallen R. Tranexamic acid as an aid to reducing blood transfusion requirements in gastric and duodenal bleeding. *BMJ* 1987;**294**:7–10.

CHAPTER 8

Cardiovascular pathophysiology

PRIYA GEORGE, JACOB GEORGE AND ALLAN D. STRUTHERS

Introduction

The cardiovascular system is a complex system designed to deliver oxygen to peripheral tissues. Oxygenated blood is delivered in an optimum fashion by the careful regulation of regional blood flow to each organ or tissue. This regulation of flow of blood is in turn dependent not only on the overall total blood volume but also on the local pressure differences within each organ. Pressure, flow and volume within the cardiovascular system are influenced in turn by both nervous impulses and circulating hormones.

In a young healthy adult, at rest, each ventricle fills during diastole to reach an end-diastolic volume of 120 mL of blood. During the following systole, 70 mL of blood is ejected [i.e. the stroke volume (SV)]. Unlike systemic organs, coronary perfusion, however, occurs during diastole. The normal ejection fraction is therefore of the order of 58%, but can lie anywhere between 55% and 75%. Overall, cardiac output (CO) is determined by two factors: SV and heart rate (HR):

$$\mathrm{CO} = \mathrm{SV} \times \mathrm{HR}$$

Starling's law states that, in a healthy heart, the force of contraction increases as diastolic stretch of the heart is increased, i.e. the more the ventricle fills with blood, the more blood will be expelled at each beat; thus, the ejection fraction is relatively constant but the absolute volume of blood ejected increases as the end-diastolic volume increases. The control of the CO is, however, more complex than just Starling's law. CO is, in fact, determined by several variables:

- *Preload*, i.e. venous filling. This influences CO by Starling's law (see above). The two main determinants of preload are the circulating blood volume and venous tone. In a normal subject, a reduced preload will reduce the CO, but this is not true in myocardial disease (see below).
- *Cardiac contractility*. The optimal sarcomere length is 2.2 μm. Stretching of the sarcomere will increase the number of actin–myosin bridges and therefore result in an increased generation of force.
- *Afterload*, i.e. arterial resistance. Arterial resistance acts as an impediment to cardiac ejection. The main contributors to arterial resistance are the arterioles and not the large arteries. A reduced afterload due to an arterial vasodilator will normally increase the CO because it reduces the restraining effect of arterial stiffness.
- *HR*. This obviously influences CO, since CO = SV × HR.

In turn, these three components (CO, SV and HR) are influenced by several regulatory mechanisms:

- Sympathetic nervous activity: the sympathetic nervous system causes norepinephrine (noradrenaline) to be released within synaptic clefts, which then increases HR, cardiac contractility, preload and afterload.
- Circulating catecholamines.
- Parasympathetic nervous system.

Circulating catecholamines

Epinephrine (adrenaline) is a circulating hormone released by the adrenal medulla at times of extreme stress. It is mainly a β-agonist with some α-agonist activity. Its β_1 activity increases HR and cardiac contractility, whereas its β_2 activity causes peripheral vasodilation of both arteries and veins, hence causing decreased afterload and decreased preload.

Norepinephrine is very different from epinephrine in that norepinephrine is principally the neurotransmitter for the sympathetic nervous system. It mainly acts locally in tissues but it does spill over from the synaptic clefts to appear in the bloodstream. Although norepinephrine is mainly a neurotransmitter, the 'spilled over' part of norepinephrine does circulate in the bloodstream and act as a circulating hormone.

Norepinephrine is mainly an α-agonist with a little β_1-agonist activity. Its α-agonist effect leads to peripheral vasoconstriction of both arteries and veins, and therefore increased preload and afterload, whereas its β_1 activity increases cardiac contractility and HR. It is worth noting that epinephrine decreases preload and afterload whereas norepinephrine increases preload and afterload. Both substances increases HR and cardiac contractility.

Parasympathetic nervous system

This regulates cardiac function through the vagus (X) nerve. The principal effect of vagal stimulation is to reduce the HR, partly by slowing conduction at the atrioventricular (AV) node. Indeed, the intrinsic HR of a young adult is about 110 beats/min but the tonic influence of the vagus reduces this to a normal resting rate of 70 beats/min. The vagus has little effect on preload, afterload or force of contraction.

Peripheral circulation

The large arteries passively channel the blood ejected from the left ventricle to the peripheries. To designate this, these vessels are normally called *conduit vessels*. On the other hand, the arterioles have a muscular wall that is thick by comparison with their lumen and they make the largest contribution to afterload, which is why they are called the *resistance vessels* of the body. It is likely that at any one time some arterioles are open and some closed. The open arterioles offer little resistance. Therefore, peripheral resistance is mainly determined by the number of arterioles which are closed at any one time.

Capillaries, like arterioles, are in the state of being either open or closed. For example, in skeletal muscle at rest, only two or three capillaries are open in each mm^3 of tissue, whereas this goes up to 200–300 during exercise. In the skin and other tissues, there are numerous arteriovenous anastomoses that allow a variable amount of blood to bypass the capillary bed altogether. The venous bed, especially the venules of the liver and spleen, act as a reservoir for the majority of the circulating blood volume. To designate this, the peripheral veins are normally called the *capacitance vessels* of the body. Small veins and sinusoids contain about 45% of the intravascular volume while the large central veins hold about the 18%. Mean right atrial pressure is approximately 0 mmHg whereas the capillary pressure is about 15 mmHg, so that the return of blood to the heart through the veins is a passive process, assisted by the pumping action of skeletal muscles, which compress segments of veins and hence propel blood towards the heart.

Blood pressure

Blood pressure (BP) is a composite term involving both CO and total peripheral resistance (TPR). The reason that BP remains a commonly used parameter is really to do with the fact that, in any given patient, BP is easily measured repeatedly and by non-invasive means, and also because a chronically elevated BP has prognostic implications in the long term. In the case of a low BP, however, its individual components (CO and TPR) give much more useful information about the state of the cardiovascular system but their measurement is more complex and more invasive. For example, shock can usually be defined as a low BP but this in itself is little guide to the cause or the treatment required. Knowledge of the CO and TPR would help enormously since cardiogenic shock is characterized by a very low CO (with a normal or high compensatory TPR) whereas septicaemic shock is characterized by a low TPR (with a normal or high compensatory CO).

Hypertension

Hypertension is defined as an elevated systemic BP. The current National Institute for Health and Clinical Excellence guidelines define hypertension as a BP measurement of systolic, diastolic or both above 140/90 mmHg on two separate occasions. The threshold for commencing treatment is if the patient either has a BP of 160/100 mmHg or greater, *or* has a BP of 140/90 mmHg *and* one of the following two criteria (1) a 10 year cardiovascular disease risk >20% or (2) existing cardiovascular disease/target organ damage. It appears, however, that the cardiovascular system changes as time goes on in a patient with hypertension. In the early stages of essential hypertension, the CO is increased while the TPR is normal. After an undefined period of years, the CO returns to normal but the TPR rises. Indeed, an increased TPR is the main cardiovascular hallmark of established essential hypertension.

The vast majority of hypertension is chronic and not immediately life-threatening, but it must be appreciated that there is a form of disease called accelerated or malignant phase hypertension. Often, malignant phase hypertension develops in a patient who has for many years had benign, albeit poorly controlled, hypertension. The factors that trigger benign hypertension to change into malignant hypertension are unknown. Malignant hypertension is characterized pathologically by fibrinoid necrosis in small arteries and arterioles, this being particularly obvious in the renal arterioles. Clinically, malignant phase hypertension presents as a grade 3 retinopathy (haemorrhages and exudates) or a grade 4 retinopathy (papilloedema) in a patient with a markedly elevated BP (diastolic BP >120 mmHg). Patients usually also have proteinuria and an elevated plasma creatinine, although neither of these are diagnostic as they can occur in the late stages of benign hypertension. The prognosis of malignant phase hypertension was extremely poor (median survival 6 months) in the 1950s before antihypertensive therapy was available. Nowadays, malignant phase hypertension is treated by immediate hospital admission and the institution of oral antihypertensive medication with the aim of reducing the BP into the normal range over the course of the next 3–4 days. Parenteral antihypertensive therapy is not usually indicated but may be used if patients are ‘nil by mouth’ preoperatively or have impaired swallowing reflexes.

The vast majority (approximately 90%) of patients have essential hypertension, in which there is no known cause for the elevated BP. When there is known cause, this is called secondary hypertension. These are various causes for secondary hypertension and these will be dealt with in turn. Secondary hypertension represents a minority of hypertensive patients, but they are a group of particular importance because surgery may play a crucial role in their management.

Hypertension and surgery

Pre-existing hypertension is the most common medical reason for delaying surgery. The hypertensive patient is obviously at greater risk during surgery than the normotensive patient. For example, the risk of postoperative myocardial infarction (MI) is doubled in a hypertensive patient. These increased haemodynamic risks of MI and ischaemia are reduced if the BP is treated prior to surgery. If diastolic BP is >110 mmHg, then it is wise, if possible, to postpone surgery until, ideally, diastolic BP is 85–95 mmHg. In a study by Goldman and Caldera, patients with diastolic BP <110 mmHg did not appear to be at significantly higher operative risk. The significance of systolic hypertension was not commented upon but, in practice, values >160 mmHg are often taken as a reason to delay surgery.

Some consider that regional anaesthesia is safer than general anaesthesia in a hypertensive patient, but epidural anaesthesia can itself cause marked vasodilation and BP swings, which are themselves poorly tolerated in untreated hypertensives. If regional anaesthesia is employed, then careful monitoring and careful attention to fluid balance are essential.

In the postoperative period, BP can be particularly unstable if there is a prolonged hypotensive effect of anaesthesia owing to impaired renal or hepatic clearance, blood loss, sepsis or pain. Another important cause of labile (usually low) BP postoperatively is the development of atrial fibrillation with uncontrolled ventricular rate (so-called 'fast AF'). As 20% of CO is dependent on atrial contraction ('atrial kick'), the loss of this, especially in patients with heart failure, can result in hypotension, breathlessness and compensatory sympathetic activation. Inadequately controlled pain or bladder distension will cause sympathetic activation that can elevate the BP and cause tachycardia. Both of these will increase myocardial O_2 consumption, which may precipitate myocardial ischaemia or infarction. Therefore, it is sensible to monitor and control both BP and HR in the postoperative period.

Therapeutic management of hypertension

Most oral antihypertensive agents can be taken with small sips of water on the morning of surgery. The therapeutic options available for parenteral use at this time are sodium nitroprusside (SNP), hydralazine, nitrates, dihydropyridine calcium channel blockers, beta-blockers, fenoldopam or enalaprilat (last two not available in UK). The first two have the disadvantage that they can both cause a reflex tachycardia so that they may need to be combined with a cardioselective beta-blocker (e.g. metoprolol 5 mg i.v.) if tachycardia becomes troublesome. SNP is a potent arterial and venous vasodilator. Cyanide toxicity can also occur with prolonged use (>24 hours) of SNP, especially in patients with vitamin B_{12} deficiency. B_{12} is a cofactor for the mitochondrial enzyme rhodanase that catalyses the conversion of cyanide to the relatively harmless thiocyanate. Intravenous glyceryl trinitrate (GTN) has the advantage of similar action to SNP (but predominantly venodilation) without the risk of cyanide toxicity. It is particularly useful in patients with known angina.

In the UK, the only dihydropyridine calcium channel blocker available in intravenous preparation is nimodipine. It is reserved for the treatment of intracranial hypertension only. In other parts of the world, both clevidipine and nicardipine are available. Clevidipine is very short acting but, unlike other dihydropyridine calcium channel blockers and GTN, does not cause reflex tachycardia. Nicardipine and enalaprilat are long-acting agents and therefore may be of limited use because of the risks should the patient have hypotension secondary to other causes such as bleeding or sepsis.

It is also crucial in the preoperative and postoperative phases that all regular medication be continued. This is especially true of antihypertensive or antianginal medication, for which parenteral administration may be required, e.g. atenolol 4 mg/h by continuous i.v. infusion. Similarly, antiarrhythmic therapy must be continued by parenteral administration.

Secondary hypertension

Renovascular hypertension (renal artery stenosis)

Renovascular disease or renal artery stenosis can be either unilateral or bilateral. It can be caused in young, usually female, patients by fibromuscular dysplasia or in older patients of either sex by atheromatous disease. It should be suspected in patients younger than 30 years, in those with treatment-resistant (more than three agents) hypertension, in patients with known peripheral vascular disease with sudden worsening of control, diabetic patients or patients on angiotensin-converting enzyme (ACE) inhibitors/angiotensin receptor blockers whose renal function deteriorates (>20% worsening) on initiation of therapy or dose uptitration. An initial renal ultrasound scan is appropriate to determine renal structure and cortical thinning. Further investigations can be done by MAG-3 (radioisotope imaging) or magnetic resonance renal angiogram. However, the largest clinical trial to date, the Angioplasty and Stenting for Renal Artery Lesions (ASTRAL) trial, showed no difference in all significant endpoints between patients who underwent stenting and the control group. There are many reasons that this may be so, not least because the population studied was heterogeneous. A meta-analysis by Leertouwer *et al.* showed that the cure rate from stenting was approximately 20%, and 30% of patients also experienced an improvement in their renal function.

The crucial point about renovascular disease is that BP may unfortunately remain high in some patients even after surgical correction. The reason for this is unclear, but it may be because the patient had underlying essential hypertension to begin with or because, once established, an elevated BP may perpetuate itself. Because surgery may not always be curative in all patients, a great deal of effort has gone into trying to predict the BP response after surgery in individual patients. Several tests have been suggested, such as measuring the renal vein renin ratio or the renin response to an ACE inhibitor. Both tests give some guide but are not reliable enough to be used alone in individual patients. The decision on whether to attempt correction is therefore usually based on a composite of several clinical features such as:

- age
- level of BP on medical therapy
- rate of renal function decline
- hemodynamic significance of the lesion
- patient preference.

In other words, each patient should be assessed on an individual basis.

The choice of treatment lies between one of three options:

1 intervention by percutaneous transluminal angioplasty (PTA)
2 medical therapy but excluding an ACE inhibitor
3 medical therapy including an ACE inhibitor.

In fibromuscular dysplasia in young subjects, intervention by surgery or PTA is usually advocated. In atheromatous disease, medical therapy is more standard with consideration of intervention if BP is uncontrolled or if overall renal function deteriorates. As for medical therapy, it is normal to include a β receptor antagonist (beta-blocker) as this helps lower the normally elevated plasma renin activity. ACE inhibitors may well help control BP effectively but they have a high likelihood of worsening overall renal function. This is because, in the ischaemic kidney, angiotensin II (ATII) levels are high in order to constrict the glomerular efferent arteriole and maintain the glomerular filtration rate (GFR). An ACE inhibitor is likely to remove the protective effect of ATII and cause the GFR to fall dramatically. If this occurs in bilateral disease, plasma creatinine will rise; if this occurs in unilateral disease, the plasma creatinine may well stay normal owing to the other kidney functioning, although the drug will effectively have produced a 'medically unilateral nephrectomy'. In short, the physician may try an ACE inhibitor but only if the renal function is carefully monitored. The patient will not thank the doctor for having a normal BP at the cost of acute kidney injury.

Primary hyperaldosteronism

This will be dealt with fully in Chapter 20. The main problem is in distinguishing a single adrenal adenoma from bilateral adrenal hyperplasia. The gold standard for making this differentiation is selective adrenal vein sampling with measurement of the aldosterone/cortisol ratio in each vein but this is a very invasive procedure. This is often not required as CT scanning and/or ^{131}I-iodocholesterol scans with dexamethasone suppression will often distinguish between single adenomas and bilaterally nodular hyperplasia.

Medical therapy with spironolactone is preferred in bilateral hyperplasia. In those with a single adenoma, preoperative spironolactone for 3–4 weeks helps to control the BP prior to surgery. Surgery is curative only in those patients who have unilateral disease with a success rate of up to 60%. Laparoscopic adrenalectomy is the procedure of choice since it is associated with shorter recovery times and fewer complications than open procedures.

Phaeochromocytoma

This is dealt with in more detail in Chapter 20. The classic triad of symptoms includes episodic headaches, tachycardia and sweating. Less common symptoms include weight loss, polyuria and postural hypotension. It is worth emphasizing that, although a rare diagnosis, it is a crucial diagnosis to make since 10% are malignant. It is also worth bearing in mind that phaeochromocytomas may be part of a multiple endocrine neoplasia syndrome type 2.

Phaeochromocytoma is screened for by measuring at least two 24 hour urinary free catecholamines collections, plasma or urinary metanephrines or plasma chromogranin A. These biochemical tests should also be carried out in patients who have incidental adrenal masses detected on CT/MRI. Preoperative localization involves CT or MRI. If there is high biochemical suspicion but negative CT/MRI, functional localization with an MIBG (metaiodobenzylguanidine) scan or ^{111}In-pentetreotide scintigraphy nuclear scanning and/or venous sampling along the whole length of the superior and inferior venae cavae with samples from most of their tributaries can be considered.

Preoperative treatment should first aim to replete the contracted intravascular volume with saline. Hypovolaemia and sodium excretion is catecholamine driven. Pharmacological treatment with an alpha-blocker *before* beta-blockade is crucial. The alpha-blocker of choice is phenoxybenzamine (initially 10 mg b.d.). The dose should be increased to produce mild postural hypotension and it should be given for 1–3 weeks before surgery. More selective alpha-blockers such as doxazosin or prazosin may be used but are less effective than phenoxybenzamine. The beta-blocker must not be given without the phenoxybenzamine since endogenous catecholamines will then stimulate α receptors to cause profound vasoconstriction without the restraining influence of coincidental vasodilatory β_2 receptor stimulation. Despite this careful preparation, preoperative swings in BP are common, often requiring intravenous fluid or intravenous SNP.

The electrocardiogram

The contraction of any muscle is associated with electrical changes called depolarization and these changes can be detected by electrodes attached to the surface of the body.

The two atria contract together to produce the P wave, which is therefore due to atrial depolarization; the QRS complex represents rapid depolarization of the ventricles, and the subsequent T wave is the result of ventricular repolarization.

While recording an ECG, the speed at which the paper runs is constant (25 mm/s) so that the HR can be determined by counting the number of large squares on the ECG paper between one QRS complex and the next QRS complex. The HR in beats/min is calculated by dividing the number 300 by the number of large squares on the ECG paper that separate two consecutive QRS complexes. A simple trick to determine the exact rhythm in a tachycardic patient is to increase the speed of the ECG acquisition to 50 mm/s. This has the effect of 'slowing' the trace to enable the visualization of ECG complexes.

An ECG is not just one view of the heart rhythm. It is, in fact, six vector views (I, II, III, aVR, aVL, aVF; Figure 8.1) and six surface views (V_{1-6}) of the heart. The 12 views coming from each of the 10 leads of the standard ECG. The upward (R wave) deflection in leads V_1 and V_2 reflects the electrical activity of the right ventricle while the upward (R wave) deflection in leads V_5 and V_6 reflects the electrical activity of the left ventricle. Similarly the downward (S wave) deflection in V_1 and V_2 reflects the left ventricle while the downward deflection in V_5 and V_6

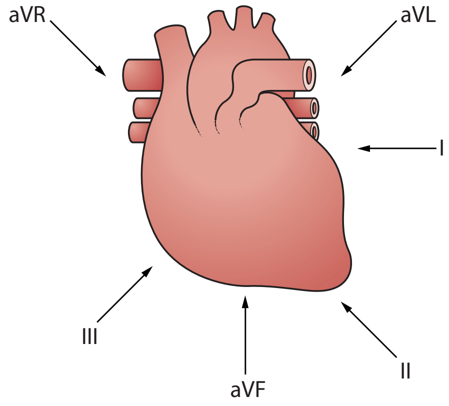


Figure 8.1 Diagrammatic representation of the angle from which each of the limb leads views the heart

reflects the right ventricle. Hence, an enlarged left ventricle will produce a tall R wave in V_6 and a large deep S wave in V_1. The ECG criterion for left ventricular hypertrophy is therefore:

$RV_6 + SV_1 > 35$ mm (i.e. seven large vertical boxes on ECG paper)

Similarly, right ventricular (RV) hypertrophy is diagnosed by an R wave in V_1 that is so tall as to be larger than the S wave, which is normally predominant in V_1, i.e. in RV hypertrophy:

$$RV_1 > SV_1 \text{ or } R/S \text{ in } V_1 > 1$$

Each limb lead views the heart from a certain position: aVR views the heart from the right arm, aVL from the left arm and aVF from the floor. Therefore, the inferior surface of the heart is best seen by II, III and aVF, so that an inferior MI produces ECG changes which are mainly or exclusively seen in leads II, III and aVF. By contrast, an anterolateral MI is mainly or exclusively seen in leads II, III and aVF. By contrast, an anterolateral MI affects mainly the left ventricle so that the ECG changes will be mainly seen in I, aVL, V_5 and V_6 since these are the leads that produce the best view of the left ventricle.

Arrhythmias

In any abnormal rhythm, the first point to note is the shape and duration of the QRS complex. If it is narrow and lasts for <0.12 seconds (three small squares or 0.04×3), the rhythm is supraventricular in origin (this includes sinus rhythm and other supraventricular rhythms). If the QRS is broad and lasts for >0.12 seconds, the rhythm is either ventricular in origin or supraventricular with bundle branch block of some kind. Ventricular rhythms have bizarre and broad QRS complexes. Apart from ventricular extrasystoles, other ventricular rhythms are ventricular tachycardia, idioventricular rhythm or ventricular fibrillation. Ventricular tachycardia is at least three ventricular extrasystoles in a row when the rate is >100 beats/min. Idioventricular rhythm is at least three ventricular extrasystoles in a row when the rate is <100 beats/min; both, but especially the latter, can occasionally be well tolerated and produce little haemodynamic upset. Ventricular fibrillation is chaotic irregular electrical activity in which no formed P, QRS or T wave complexes are discernible and in which the patient invariably becomes unconscious owing to loss of cardiac output and therefore a cardiac arrest has occurred.

There are many different types of supraventricular rhythms. The commonest is atrial fibrillation, in which the baseline ECG is irregular with no P waves and narrow normal looking QRS complexes occurring at irregular intervals. Atrial flutter is similar but the baseline has a saw tooth appearance with regular flutter waves instead of true P waves; in atrial flutter, the shape of the QRS complexes is normal although they may occur either regularly or irregularly. Sometimes it is not possible to differentiate in a given patient between atrial fibrillation and atrial flutter so that the term atrial flutter/fibrillation is used. Since the treatment (with digoxin, beta-blocker, verapamil, diltiazem or DC shock) is the same, it is not crucial to differentiate flutter from fibrillation. The third group of abnormal supraventricular rhythms are generally called supraventricular tachycardia (SVT). SVT can be further divided into atrial tachycardia or junctional (= nodal) tachycardia, but this distinction is fairly academic as treatment is the same for both types of SVT. In both kinds of SVT, the QRS complexes are narrow, normal in shape and the HR is ≥150 beats/min.

The HR itself is a useful clue as to whether a tachycardia is just a sinus tachycardia or SVT. The former are usually <135 beats/min whereas the latter are usually >145 beats/min. Clinically this is an important distinction to make as heart failure often causes sinus tachycardia and treatment should be directed to the heart failure and not specifically to the sinus tachycardia. By contrast, SVT often causes a degree of heart failure and treatment should be directed initially against the SVT rather than the heart failure.

Heart block

In first-degree heart block, the PR interval is greater than 0.2 seconds (or one large square).

Second-degree heart block has several subtypes. Second-degree heart block is usually below the level of the AV node/His bundle and therefore categorized as 'infra-Hissian'. It is divided into the following broad categories:

- *Mobitz type I (Wenkebach type):* there is progressive lengthening of the PR interval followed by a P wave with no subsequent QRS.
- *Mobitz type II:* fixed prolonged PR interval with one isolated P wave is not followed by any QRS.
- *A regular pattern* (e.g. 2:1 or 3:1) in which every second or third P wave has no subsequent QRS complex.

In third-degree (or complex) heart block, the P waves and the QRS complexes occur independently and at different times from each other.

Heart block can also occur in the right and left bundle branches. In bundle branch block, the QRS complex is widened (>0.12 seconds or more than three small squares). Right bundle branch block is best seen in V_1 in which there is an RSR pattern to the QRS complex. Left bundle branch block is best seen in V_6 in which the QRS complex forms an M pattern. In both cases, the subsequent T wave is inverted.

In MI, there are four characteristic ECG changes:

1. hyperacute T waves (early on)
2. ST elevation
3. T wave inversion
4. Q waves (later on).

These changes are seen mainly in II, III and aVF in inferior MIs and in I, aVL, V_4–V_6 in anterolateral MIs. Other leads may show reciprocal changes (i.e. marked T wave inversion only). One subgroup of MIs are non-ST elevation MIs (NSTEMI), in which the infarcted muscle is less extensive. In NSTEMIs, either no Q waves are found but very prominent T wave inversion occurs or new left bundle branch block occurs. Surgery should preferably be delayed for at least 3 months after any kind of MI. This is not the case after an episode of angina so that it is often important to ascertain whether an episode of chest pain has been ischaemia or infarction as confirmed from troponin T (or I) levels and clinical diagnosis.

An elevated ST segment is due either to an MI or to pericarditis. In the former, the ST elevation is usually convex in shape and limited to certain ECG leads. In pericarditis, the ST elevation is concave and is widespread over many ECG leads. In contrast, ST depression often coincides with T wave inversion and this combined pattern occurs in many situations, e.g. in the presence of ischaemia, bundle branch block, ventricular hypertrophy ('strain pattern'), hypokalaemia or in any patient receiving digoxin therapy.

Pulmonary embolism

The ECG changes seen in pulmonary embolism (PE) are:

- sinus tachycardia (most common)
- T wave inversion V_{1-3} (greatest sensitivity for diagnosis of PE)
- right bundle branch block
- SI QIII TIII is a sign of acute cor pulmonale (sensitivity of 54% and a specificity of 62% often quoted, and also been reported in air embolism and pneumothorax; see Ferrari *et al.*)
- right ventricular hypertrophy and strain (i.e. tall R in V_1, tall P waves).

These changes can be used as tentative support only to a clinical suspicion of PE. The clinician should confirm the diagnosis of PE by either a *V/Q* scan (if the patient has a normal chest radiograph) or a CT pulmonary angiogram before long-term anticoagulation is begun. Pregnant patients should undergo a perfusion scan only to minimize fetal radiation exposure.

Cardiac arrest

Cardiac or circulatory arrest can occur under numerous and varied circumstances, which makes it inappropriate to define a rigid and didactic strategy for its management. The correct approach to cardiopulmonary resuscitation (CPR) is therefore a flexible one that can then be tailored to individual circumstances.

There is a wide differential for causes of fairly sudden loss of consciousness, including vasovagal syncope, cerebrovascular accident, epilepsy, hypoglycaemia, drug overdose or trauma. These should be considered if both a carotid pulse and respiration are still present in a suddenly unconscious patient. If on initial assessment a carotid pulse is palpable but respiration absent, then this is respiratory arrest and attention should be focused on maintaining a good airway and artificial ventilation. If no carotid pulse is present then cardiac arrest can be assumed to have occurred and CPR should be undertaken.

Types of cardiorespiratory arrest

Ventricular fibrillation/pulseless ventricular tachycardia

Causes include:

- ischaemic heart disease (usually post-MI)
- structural heart disease (e.g. hypertrophic cardiomyopathy/dilated cardiomyopathy)
- severe electrolyte abnormalities
- drug-induced/congenital long QT interval.

Pulseless electrical activity/electromechanical dissociation: electrical activity on ECG but no pulse

Causes include:

- hypothermia
- hypoxia
- hypo-/hyperkalaemia
- hypovolaemia
- tension pneumothorax
- PE
- cardiac tamponade
- toxic/overdose.

Asystole

Respiratory arrest alone (carotid pulse present)

Causes include:

- respiratory disease, e.g. foreign body, pneumothorax
- respiratory depressant drug (e.g. intravenous diazepam for a minor surgical or dental procedure).

Initially, the patient should be positioned flat on the floor or on a firm surface. The head should be extended and the airway inspected so that any foreign matter can be removed directly from the larynx. Artificial ventilation should be begun using a Guedel airway, a mask and a respirator bag. As soon as personnel are available, the patient should be intubated and ventilated with 100% oxygen. It must be checked by stethoscope that both lungs are being ventilated equally in order not to overlook a pneumothorax, a lodged foreign body or an endotracheal tube that has been pushed past the tracheal bifurcation. The latest Resuscitation Council guidelines are shown in Figure 8.2.

At this stage, as well as obtaining an ECG, intravenous access should be secured. This should ideally be via a jugular, subclavian or femoral vein but an antecubital vein is better than none.

If resuscitation is successful, it is crucial to review possible reversible causes to prevent repetition of the arrest. Blood should be assessed for full blood count, glucose, electrolytes, arterial pH and gases. A full 12-lead ECG should be taken to exclude MI, arrhythmia, heart block or PE. A chest radiograph should also be taken to exclude pneumothorax, aspiration, tamponade, aortic dissection and to check the position of any endotracheal tube.

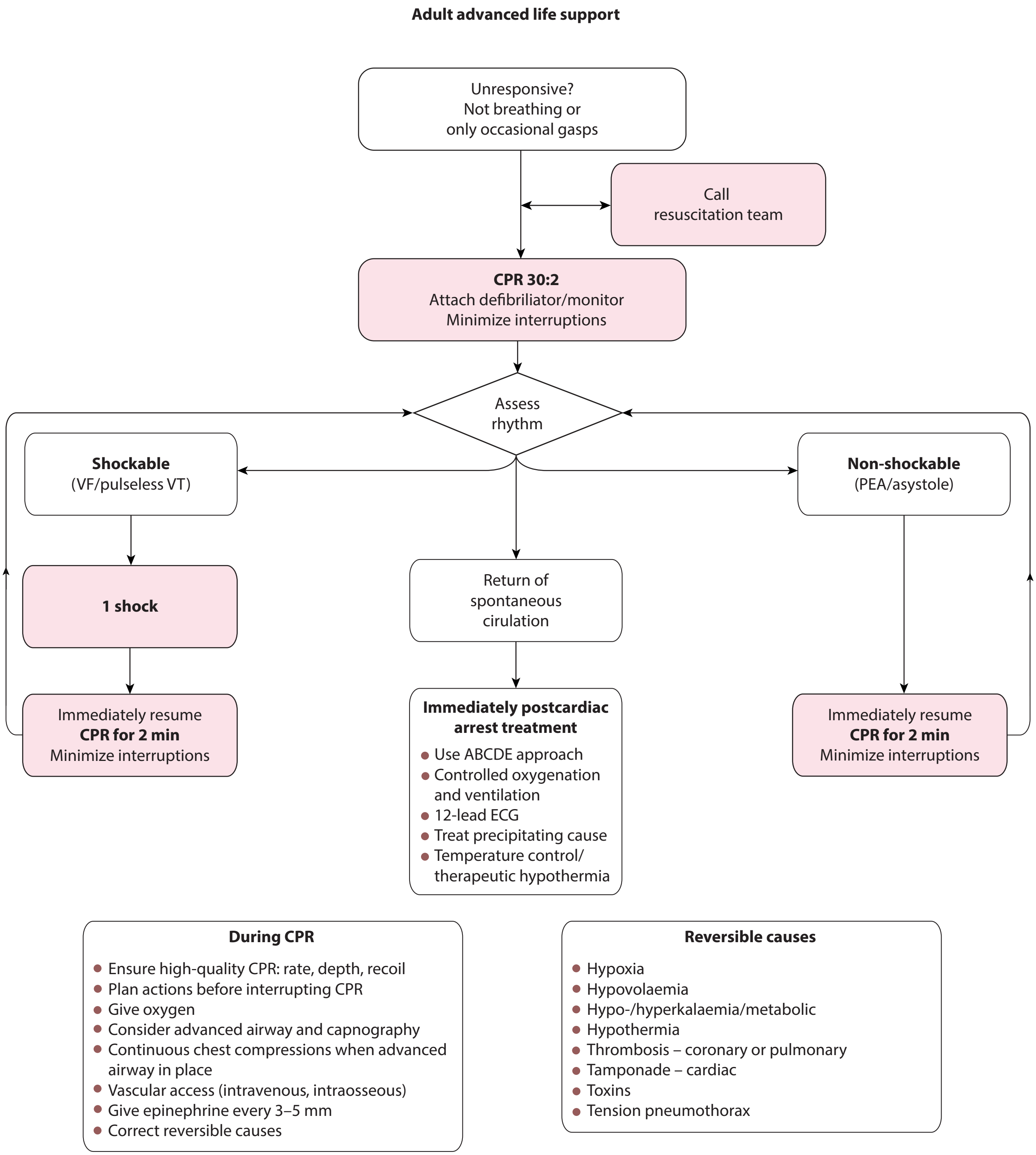

Adult advanced life support
Unresponsive? Not breathing or only occasional gasps
Call resuscitation team
CPR 30:2
Attach defibriliator/monitor
Minimize interruptions
Assess rhythm
Shockable
(VF/pulseless VT)
1 shock
Immediately resume
CPR for 2 min
Minimize interruptions
Return of spontaneous cirulation
Immediately postcardiac arrest treatment
Use ABCDE approach
Controlled oxygenation and ventilation
12-lead ECG
Treat precipitating cause
Temperature control/ therapeutic hypothermia
Non-shockable
(PEA/asystole)
Immediately resume
CPR for 2 min
Minimize interruptions
During CPR
Ensure high-quality CPR: rate, depth, recoil
Plan actions before interrupting CPR
Give oxygen
Consider advanced airway and capnography
Continuous chest compressions when advanced airway in place
Vascular access (intravenous, intraosseous)
Give epinephrine every 3–5 mm
Correct reversible causes
Reversible causes
Hypoxia
Hypovolaemia
Hypo-/hyperkalaemia/metabolic
Hypothermia
Thrombosis – coronary or pulmonary
Tamponade – cardiac
Toxins
Tension pneumothorax

Figure 8.2 Current Resuscitation Council UK advanced life support algorithm. CPR, cardiopulmonary resuscitation; PEA, pulseless electrical activity; VF, ventricular fibrillation; VT, ventricular tachycardia. (Reproduced with the kind permission of the Resuscitation Council UK.)

Cardiac failure

Pathophysiology

This is defined as an inability of the heart to meet tissue requirements for oxygen. In the vast majority of cases this is due to myocardial disease causing a low CO (low output cardiac failure). In a minority of cases, this is because of an increase in tissue requirements for oxygen such as thyrotoxicosis, anaemia, AV malformation or Paget disease of the bone – the so-called ‘high output failures’. In these cases, the CO is increased but is still insufficient for tissue requirements.

Heart failure is divided into right- and left-sided heart failure depending on the relative degrees of impairment of each ventricle. Classification can also be based on whether the impairment is in systolic or diastolic (also known as heart failure with normal ejection fraction) function. The commonest cause of left heart failure is ischaemic heart disease leading to MI that makes part of the left ventricle non-contractile. Cardiomyopathy or cor pulmonale are other causes of heart failure but the medical treatment of all three of the above forms of heart failure are similar so that it is not quite so crucial to differentiate them. Almost all the evidence base for therapy in heart failure comes from trials in systolic dysfunction. At present, there are no therapies that have been proven to improve survival in diastolic heart failure and therefore these patients are managed primarily for symptom control with diuretics and treatment of other cardiovascular risk factors such as hypertension.

It is therefore mandatory to think what has caused heart failure in any given patient since the treatment is very different depending on the underlying causes. Too often, for example, it is not noticed that the cause of heart failure in a given patient is the arrhythmia of uncontrolled atrial fibrillation. Diuretics alone do little to help such patients whereas bringing the ventricle rate down to 80 beats/min from 140 beats/min by the correct use of digoxin, beta-blockers or DC cardioversion will effect a more meaningful change in their cardiac function. An ECG and echocardiogram is essential in every dyspnoeic patient to establish the heart rhythm. Fast atrial fibrillation is often precipitated in a surgical patient by a postoperative chest infection.

It is essential to remember that a patient may have more than one cause for heart failure. For example, a patient with mitral valve disease will often have fast atrial fibrillation, the substrate being an enlarged left atrium. Similarly, an older patient with aortic valve disease may in addition have had a previous MI leading to poor ventricular function. If the latter disease is more severe than the former, then aortic valve surgery may have little impact on symptoms and the correct procedure would be optimization of medical therapy for the ischaemic heart disease. All patients with heart failure therefore require to be thoroughly assessed for the underlying cause or causes: if more than one cause exists, then the relative contribution of each cause must be thoroughly assessed in order to plan the correct treatment in that individual.

Whatever its aetiology, one of the clinical hallmarks of heart failure is renal sodium and water retention, which ultimately causes ankle oedema, raised jugular venous pressure and pulmonary oedema. Sodium (and therefore fluid) retention occurs as a maladaptive compensatory process in response to many pathophysiological changes seen in heart failure. For example, reduced tissue perfusion from forward pump failure and reduced intravascular volume from poor backward negative ‘suction’ pressure from the right ventricular dysfunction. This can easily be inferred from the fact that a major part of the treatment of heart failure is diuretic therapy, which is designed to reverse this sodium and water retention.

Principles of treatment

It is crucial to appreciate that the aims of therapy in acute and chronic heart failure are very different with acute cardiac decompensation being more relevant to surgical practice. In the acute case, the aim is often to support cardiovascular function until it recovers spontaneously as a result of the natural healing processes which occur after surgery (or even after MI). This is achieved by consideration of ventricular function curves and the correct use of combined therapy with diuretics, inotropic drugs and vasodilators, in order to increase systemic BP and SV while also decreasing filling pressures. It is felt that optimizing these three haemodynamic parameters keeps patients alive longer so that spontaneous healing is given a longer timescale in which to operate. Chronic heart failure is very different in that optimizing haemodynamic parameters would seem to be a logical step but, surprisingly, there is little correlation between the acute haemodynamic effects of various drugs and their long-term effects on morbidity and mortality so that the acute haemodynamic effects of drugs are of lesser importance. Invasive haemodynamic monitoring is therefore of relevance only to the acutely ill cardiac patient.

The Starling curve is a physiological mechanism that occurs in normal man. In heart failure, the pattern is rather different: the ventricular function curve lies below the normal curve. In addition, the curve is much flatter so that the large increase in end-diastolic pressure produces only a small, if any, increase in SV (Figure 8.3).

It is important to realize that there are clinical implications to this ventricular function curve (Figure 8.4). If this renal sodium/water retention overshoots, then the end-diastolic pressure rises excessively and pulmonary oedema occurs.

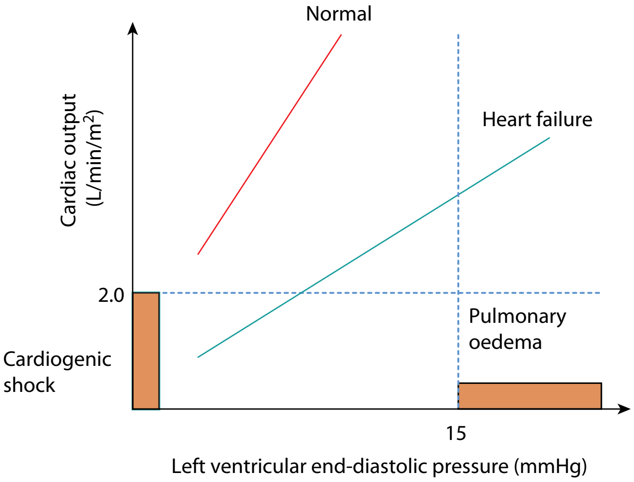


Figure 8.3 The ventricular function (or Starling) curve for a normal heart and for a patient with heart failure showing when cardiogenic shock and pulmonary oedema occur.

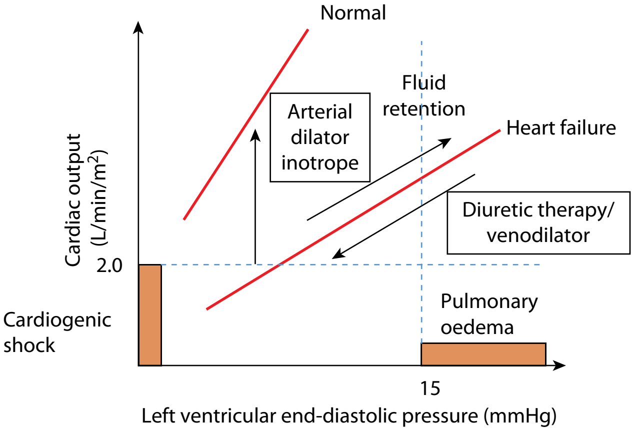

Normal
Fluid retention
Arterial dilator inotrope
Heart failure
Diuretic therapy/ venodilator
Cardiac output (L/min/m^2)
2.0
Pulmonary oedema
Cardiogenic shock
15
Left ventricular end-diastolic pressure (mmHg)

Figure 8.4 The effect of various drugs on ventricular function.

In older terminology, this is called 'backward failure' of the heart because fluid accumulates prior to the left ventricle, i.e. in the lungs. In this situation, venous pressure is very high, CO is moderately low but systolic BP is maintained. The correct treatment here is either intravenous diuretic therapy or venodilator therapy such as nitrates. Diuretics will reduce the circulating volume, hence reducing the end-diastolic pressure with it – the symptoms of pulmonary oedema. Nitrates will more directly reduce end-diastolic pressure and pulmonary oedema by reducing preload. The opposite clinical problem occur when SV falls excessively, then the syndrome of cardiac shock will ensue. This used to be called 'forward failure' of the heart. In this situation, CO and systolic BP are both very low while venous pressure is only moderately increased. Hence the correct treatment is to increase the SV: this could be achieved best by administering an inotropic drug such as dobutamine. In some other situations, the SV can be augmented not only by a directly acting inotropic drug but also by an arterial vasodilator drug. This is because, often, a low CO is due to resistance to flow in the arterial tree so that an arterial dilator will reduce afterload and hence indirectly boost SV. In the precise case of cardiogenic shock, this would not be as valuable a strategy since the arterial vasodilator might further reduce an already low systemic BP but it would be useful therapy in myocardial failure with a reasonable level of systemic BP. Therefore, in deciding about therapy of heart failure, there is a third parameter that must be considered – the systolic BP. The treatment of heart failure involves optimizing the systolic BP, the cardiac index and the pulmonary artery wedge pressure (PAWP), and in devising treatment in heart failure it is a useful exercise to think through what effect each administered drug will have on each of these three parameters.

Knowledge of the ventricular function curve for each individual allows medical therapy to be monitored more precisely. This involves the insertion of a Swan–Ganz catheter so that cardiac index and PAWP are directly measured in an individual patient. The insertion of a Swan–Ganz catheter is not justified in every case of heart failure but should be considered in the following situations:

- in severe heart failure not responsive to initial (e.g. diuretic) therapy
- in hypotensive shock, to establish its aetiology and to monitor treatment.

In both situations the aim is to optimize the three parameters of systolic BP, cardiac index and PAWP. The more specific aims are to reduce the PAWP pharmacologically, to prevent pulmonary oedema and/or to boost the CO and systolic BP in order to prevent cardiogenic shock. Some guidelines are suggested in Table 8.1.

The clinical situation in which the Swan–Ganz catheter is most useful is the patient with pulmonary oedema who is receiving intravenous nitrate therapy when knowledge of the PAWP and cardiac index will help the clinician to use a nitrate infusion rate that will reduce PAWP into an acceptable range without dropping systolic BP or cardiac index by too much. Invasive Swan–Ganz catheters are rarely used today as the management of acute heart failure has progressed to cardiac support using intra-aortic balloon pumps (IABP) and continuous positive airways pressure (CPAP).

Table 8.1 How to interpret invasive haemodynamic parameters

Systolic BP (mmHg)	PAWP (mmHg)	Cardiac index (L/min)	Diagnosis	Treatment
Normal	<15	2.7–3.5	Normal	Nil
<100	<15	<2.7	Hypovolaemic shock	Plasma expansion
<100	>18	<1.8	Cardiogenic shock	Inotropes
>100	>18	<1.8	Pulmonary oedema	Diuretics/ nitrates/ ultrafiltration

BP, blood pressure; PAWP, pulmonary artery wedge pressure.

IABPs are devices that are inserted percutaneously via the femoral artery and, as the name implies, are situated in the descending aorta. These devices inflate in synchrony with the ECG and the central aortic pressure waveform. The objective of the IABP is to inflate in diastole and therefore improve coronary perfusion (which occurs during diastole) and deflate at the end of systole, reducing afterload and therefore improving CO.

CPAP on the other hand is a non-invasive method of oxygenating the alveolar spaces by positive airway pressure. It is particularly useful in the setting of type I respiratory failure (hypoxia) due to left ventricular failure. The starting settings are usually 5 cmH_2O and are increased in increments of 2.5 cmH_2O up to 10 cmH_2O.

Another newer therapy in acute decompensated heart failure is ultrafiltration. This is in fact very similar to dialysis in that the free water component of the intravascular volume is removed. The UNLOAD trial showed significantly greater fluid loss and fewer heart failure hospitalizations with ultrafiltration than with intravenous diuretics. Perhaps most impressively, there were no statistically significant differences in creatinine levels either. Ultrafiltration is often limited by the patient's BP and must be discussed with a renal specialist in advance if a patient is not responding to intravenous diuretic therapy.

It must be emphasized that postoperative heart failure may not be simply due to myocardial dysfunction and it is crucial to exclude other cause, e.g. pericardial tamponade. This differentiation can be remarkably difficult as both conditions can produce the same findings on clinical examination, chest radiograph, ECG, echocardiography and even on Swan–Ganz catheterization. Bedside echocardiography is invaluable and can be used to help make this distinction clinically. In myocardial dysfunction, the low cardiac ejection function is evident but there are no specific pericardial abnormalities on the scan.

Cardiac pharmacotherapy

Inotropic drugs

Some, but not all, inotropic drugs also cause vasodilation. One major limitation here is that all inotropic drugs can cause arrhythmias, so that they tend to be reserved for acutely ill patients and to be used for as short a time period as the clinical situation permits. None of the inotropic agents in acute heart failure have been shown to improve survival from the condition.

Dobutamine

This acts as a β-agonist, principally on the β_1-adrenergic receptor to cause an inotropic and chronotropic effect. It can only be given intravenously. Peripheral vasodilation is also seen to a modest extent, but it does not specifically dilate renal arteries. The problem with administering dobutamine is twofold: dobutamine loses its effectiveness in time (tachyphylaxis) because of downregulation of the constantly stimulated β receptors, and, second, a proarrhythmic effect can occur. For these two reasons, dobutamine is used mainly as a temporary measure in states of acute cardiac decompensation (cardiogenic shock) but is not generally used routinely in chronic heart failure. In the latter condition dobutamine probably improves haemodynamics initially but leads to an earlier arrhythmic death (Table 8.1).

Dopamine

This if very different from dobutamine, although, like dobutamine, it also can only be given intravenously. At low dose, dopamine stimulates specific dopaminergic receptors and at higher doses it stimulates α and β receptors, although the α receptor stimulation predominates. Dopaminergic stimulation at low dose causes widespread vasodilation, especially of the renal arteries, which helps to reverse early acute renal failure. At higher dose, α receptor stimulation causes the opposite effect of vasoconstriction and decreased renal blood flow. High-dose dopamine will raise systemic BP owing to peripheral vasoconstriction, but this advantage is offset by the fact that high-dose dopamine will cause renal vasoconstriction and so increase the tendency to acute renal failure. Commonly, therefore, the dose of dopamine is kept as low as possible in order to improve any decline in renal function. Recent large multicentre, randomized controlled trials have demonstrated that low-dose dopamine administered to critically ill patients who are at risk of renal failure does not confer clinically significant protection from renal dysfunction and therefore this regimen is not currently used in routine clinical practice.

Phosphodiesterase inhibitors

β_1 and β_2 receptor stimulation leads to an increase in intracellular cyclic AMP (cAMP), which then causes their inotropic, chronotropic and vasodilatory effects. An alternative way of producing the same beneficial effects is to block the breakdown of intracellular cAMP. cAMP phosphodiesterase is the enzyme involved here so that a phosphodiesterase inhibitor will have similar effects to a β_1 and β_2 agonist. There are two such phosphodiesterase inhibitors: amrinone and milrinone. However, clinical trials such as OPTIME-CHF showed no benefit, and possibly some harm, from milrinone therapy.

Nesiritide

Nesiritide is a recombinant form of plasma B-type natriuretic peptide, which has natriuretic and vasodilator properties. Clinical trials have shown that, although nesiritide improves cardiac parameters, it did not improve survival and may worsen renal function in patients with acute decompensated heart failure.

Digoxin

Digoxin can be classified as an inotropic drug. It is used as an inotropic in chronic heart failure in sinus rhythm but here its usefulness is being disputed. It is little used in acute cardiac decompensation in which dopamine, dobutamine and amrinone are generally used for inotropic support before digoxin is given. The clearest indication for digoxin is in treatment of atrial fibrillation. Here it causes conduction delay at the AV node, which benefits patients with fast atrial fibrillation because it reduces the rate of the ventricular response to the fibrillating atria.

Digoxin has a very narrow therapeutic index, which means that its toxic dose is only marginally higher than its therapeutic dose. Hence, it is important to measure plasma digoxin levels and alter the dose of digoxin accordingly. Digoxin has a long half-life so that, even with normal renal function, only one-third of a digoxin dose is excreted within the first 24 hours after ingestion. Digoxin accumulates to potentially dangerous levels in the presence of even mild renal impairment so that monitoring of plasma digoxin levels is mandatory in the presence of any renal dysfunction.

Other vasodilators

Unfortunately, the term vasodilator includes different groups of drugs to different people. There is no doubt that nitrates, nitroprusside, hydralazine and minoxidil are vasodilators, and that the precise mechanisms of action of these drugs are uncertain. With other groups of drugs their mechanism of action is known, so that, although they are also vasodilators, they are often classified by their mechanism of action rather than under the general heading ‘vasodilators’. In other words, α_1-antagonists, calcium antagonists and ACE inhibitors are strictly speaking vasodilating drugs but, in practice, when the term ‘vasodilator’ is used, it sometimes does and sometimes does not include the α-antagonists, calcium antagonists or ACE inhibitors.

Nitrates dilate mainly the venous bed, whereas hydralazine and calcium antagonists are mainly arterial dilators. Many drugs such as nitroprusside, α_1-antagonists (prazosin) and ACE inhibitors produce a mixed pattern with dilatation of both the arterial and venous beds. Therefore, hydralazine will principally increase CO and tissue perfusion whereas nitrates will reduce pulmonary congestion.

These drugs can, of course, be used together to produce the desired effect on the ventricular function curve. In addition, they are often used along with inotropes to optimize therapy, e.g. dobutamine and nitroprusside.

The third factor which must be considered and which is not on the ventricular function curve is the systemic BP, which all vasodilators will tend to reduce and which is a limiting factor in the use of vasodilator therapy. This is another reason for combining inotropes and vasodilators to prevent excess BP reduction.

All vasodilators are contraindicated in obstructive valvular stenosis or hypertrophic obstructive cardiomyopathy but they are very helpful in valvular incompetence.

Invasive monitoring is desirable when vasodilators are used in acute cardiac decompensation in order to titrate the dose to prevent excessive reduction of preload or of CO and BP.

Nitrates

Nitroglycerin can be given sublingually, as a spray, as skin patches or as a buccal preparation, but these are mainly for use in chronic stable angina. Isosorbide mononitrate or dinitrate can also be given orally or intravenously and it is the latter which is most relevant to surgical patients with acute cardiac decompensation or patients with post-bypass hypertension.

One problem with all nitrates is that patients become tolerant to them very rapidly. After discontinuing a nitrate, a patient will rapidly lose this nitrate tolerance. Ideally, therefore, nitrates should be given only intermittently, e.g. for 12 hours of each day only, although this is difficult to achieve in an acutely ill patient.

Nitroprusside is ideal in this situation because of its rapid onset of action and its rapid cessation of activity.

Hydralazine

This is available for oral and intravenous use. Hydralazine has a delayed onset and offset of action so that unpredictable hypotension is not immediately reversible on stopping hydralazine. Intravenous hydralazine is of some value in heart failure following cardiac surgery. The combination of hydralazine and nitrates has been shown to be particularly useful in black patients in the African-American Heart Failure (A-HEFT) trial.

α_1-Antagonists

Doxazosin is the most commonly used α_1-antagonist but it is only available orally; therefore, it is mainly used to treat chronic hypertension or chronic heart failure. It is seldom used in acute cardiac decompensation.

The most relevant use of α_1-antagonists to the surgeon is in patients with phaeochromocytomas. Phenoxybenzamine is given orally for several weeks prior to surgery. Phentolamine is a different drug that is given intravenously when acute alpha-blockade is desired as, for example, during an acute hypertensive crisis due to a phaeochromocytoma. Although doxazosin, phentolamine and phenoxybenzamine sound alike and are all α-antagonists, their individual properties and their clinical uses are very different.

Other drugs

Calcium antagonists and ACE inhibitors are of lesser relevance to surgeons. Although there are three main calcium antagonists (nifedipine, diltiazem and verapamil), only diltiazem and verapamil produce AV block and hence help in the management of supraventricular arrhythmias.

ACE inhibitors are in widespread use in the treatment of chronic heart failure and essential hypertension. Many such patients will come to surgery for other reasons and an area of uncertainty exists here. The renin–angiotensin system is activated at times of hypovolaemia (e.g. during perioperative blood loss) to prevent the BP falling too much. ACE inhibitors will block the renin–angiotensin system and hence prevent this defence mechanism. Monitoring of BP and central venous pressure may thus be even more important if a patient is pretreated with an ACE inhibitor.

Antiarrhythmic agents

This topic is complex and of little relevance to surgeons except in the context of postoperative arrhythmias. There is a detailed Vaughan Williams classification of antiarrhythmic drugs but this classification is of pharmacological interest only and is of little therapeutic relevance. The problem with all antiarrhythmics drugs is twofold. They are nearly all cardiodepressant so that the minimum dose possible should be used in cardiac failure. Second, all antiarrhythmics drugs can actually promote arrhythmias (be proarrhythmic) in some subjects. The management of ventricular tachycardia and fibrillation is discussed in Cardiac arrest.

Postoperative atrial fibrillation should be managed by digoxin (if blood pressure is low and not in acute renal injury) or beta-blockers. Decision on anticoagulation must be made on an individual patient basis balancing the risk of a thromboembolic event with bleeding risk.

With supraventricular arrhythmias, the key issue is the rate at which the ventricle is beating. When the ventricular rate is >120 beats/min then the ventricle does not have enough time to fill and empty efficiently so that the CO is low. Often, therefore, reducing the ventricular rate by increasing the degree of block at the AV node (by digoxin, beta-blocker, verapamil or diltiazem) will improve the patient's clinical state dramatically. They might return to sinus rhythm, but often with digoxin, beta-blocker, verapamil or diltiazem they will remain in atrial fibrillation but at a satisfactory ventricular rate. In other words, these treatments aim to minimize the adverse haemodynamic effects of the arrhythmia but not to abolish the arrhythmia. If abolishing the supraventricular arrhythmia is the desired clinical goal, then the measures to consider are elective DC cardioversion, flecainide, disopyramide, amiodarone or overdrive pacing. Prior to these measures, carotid sinus massage or a Valsalva manoeuvre are worth trying but only in the specific case of supraventricular tachycardia since they have no useful effect in fast atrial fibrillation or atrial flutter.

GUIDE TO FURTHER READING

Costanzo MR, Guglin ME, Saltzberg MT, *et al.* Ultrafiltration versus intravenous diuretics for patients hospitalized for acute decompensated heart failure. *J Am Coll Cardiol* 2007;**49**:675–683.

Dix P, Howell S. Survey of cancellation rate of hypertensive patients undergoing anaesthesia and elective surgery. *Br J Anaesth* 2001;**86**:789–793.

Duncan JL 3rd, Fuhrman GM, Bolton JS, *et al.* Laparoscopic adrenalectomy is superior to an open approach to treat primary hyperaldosteronism. *Am Surg* 2000;**66**:932–935; discussion 935–936.

Ferrari E, Imbert A, Chevalier T, *et al*. The ECG in pulmonary embolism. *Chest* 1997;**111**:537–543.

Goldman L, Caldera DL. Risks of general anesthesia and elective operation in the hypertensive patient. *Anesthesiology* 1979;**50**:285–292.

Leertouwer TC, Gussenhoven EJ, Bosch JL, *et al*. Stent placement for renal arterial stenosis: where do we stand? A meta-analysis. *Radiology* 2000;**216**:78–85.

National Institute for Health and Clinical Excellence. *Hypertension: Management of Hypertension in Adults in Primary Care*. Clinical Guideline 64. London, UK: NICE, 2006.

Punukollu G, Gowda RM, Vasavada BC, Khan IA. Role of electrocardiography in identifying right ventricular dysfunction in acute pulmonary embolism. *Am J Cardiol* 2005;**96**:450–452.

Taylor AL, Ziesche S, Yancy C, *et al*. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. *N Engl J Med* 2004;**351**:2049–2057.

Wheatley K, Ives N, Gray R, *et al*. Revascularization versus medical therapy for renal-artery stenosis. *N Engl J Med* 2009;**361**:1953–1962.

Young WF. Primary aldosteronism: renaissance of a syndrome. *Clin Endocrinol (Oxf)* 2007;**66**:607–618.

CHAPTER 9

Pulmonary insufficiency

ELIZABETH STARREN, WILLIAM L. OLDFIELD AND ONN MIN KON

Anatomy

The airways consist of multiple tubes connected in both series and parallel fashion that can be classified into two sections according to function: the conducting zone and the respiratory zone. The conducting zone encompasses the oronasopharynx, larynx and trachea, which divides into the right and left main bronchi at the level of the second thoracic vertebrae. These subsequently subdivide into the lobar (secondary) bronchi, segmental (tertiary) bronchi and then terminal bronchioles (diameter <1 mm). This section of the lung is known as the conducting zone as its rigid, largely cartilaginous structure facilitates the passage of inspired air to the respiratory zone and the ciliated epithelium humidifies and filters air before gas exchange occurs. The first 16 divisions or orders of branching airways make up the minority of the lung volume and its blood supply is from the bronchial circulation that originates from the left side of the heart.

The respiratory zone refers to the respiratory bronchioles, alveoli, alveolar ducts and sacs. Branching orders 17–23 contain approximately 300 million alveoli, thus providing a vast surface area available for gas exchange. The primary function of the lung – gas exchange – occurs by passive diffusion of oxygen (O_2) and carbon dioxide (CO_2) across thin alveolar–capillary membranes. This gas–blood barrier comprises type I (squamous epithelial) pneumocytes, type II (surfactant producing) pneumocytes, elastic fibres, alveolar macrophages and networks of pulmonary capillaries. The blood supply is via the pulmonary circulation that originates from the right side of the heart. The pulmonary artery branches repeatedly alongside the course of the bronchi to eventually form the capillary networks that encompass the alveoli. Each red cell spends 0.75 seconds within the pulmonary capillaries to complete gas exchange. Blood is returned via the pulmonary veins.

Ventilation

The transport of oxygen and carbon dioxide

Gas transport between blood and tissues occurs, as in the lungs, by passive diffusion across a concentration gradient. Approximately 98% of O_2 is transported as oxyhaemoglobin (HbO_2), hence O_2 carriage is reduced in the anaemic state. The remaining 2% is carried as dissolved O_2. The O_2 dissociation curve (Figure 9.1) defines the relationship between O_2 tension and haemoglobin (Hb) saturation. A right shift occurs at increased hydrogen (H^+)

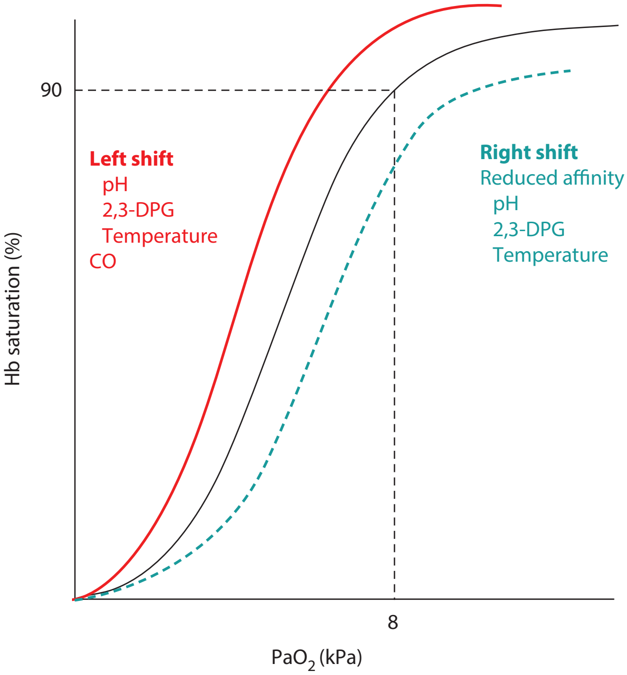


Figure 9.1 Oxygen dissociation graph. 2,3-DPG, 2,3-diphosphoglycerate; Hb, haemoglobin.

ion concentration, CO_2 tension, 2,3-diphosphoglycerate (2,3-DPG) concentration in red blood cells and body temperature, and facilitates O_2 delivery to tissues in the active, hypoxic state. A left shift occurs at reduced H^+ ion concentration, CO_2 tension and body temperature as well as in the presence of carbon monoxide. In these situations, O_2 has a higher affinity for Hb and thus unloading to tissue is reduced.

CO_2 is highly soluble and is transported within the red blood cells as dissolved CO_2, carbamino-haemoglobin (CO_2 + Hb) and bicarbonate (HCO_3). In the presence of carbonic anhydrase, carbonic acid ($CO_2 + H_2O$) is formed within the red cells and, in turn, bicarbonate is formed from the dissociation of hydrogen ions from this compound. In summary:

$$CO_2 + H_2O \leftrightarrow H_2CO_3 \leftrightarrow H^+ + HCO_3^-$$

The resultant free hydrogen ions are responsible for the Haldane effect as deoxyhaemoglobin (reduced Hb) is a greater proton acceptor than oxyhaemoglobin. In the presence of deoxyhaemoglobin (H + Hb), which is abundant in peripheral tissues, the equations are shifted right and CO_2 carriage is increased. In the presence of oxyhaemoglobin (HbO_2), as at the alveolar–capillary membrane, CO_2 is displaced and thus eliminated. In summary:

$$H^+ + HbO_2 \leftrightarrow H^+Hb + O_2$$

The pH of blood is maintained around 7.4 and small deviations in this value of blood pH can lead to severe metabolic consequences. Blood buffering is therefore extremely important in maintaining this level. Although blood contains numerous cations (e.g. Na^+, K^+, Ca^{2+} and Mg^{2+}) and anions (e.g. Cl^-, PO_4^{3-} and SO_4^{2-}) that can play a role in buffering, the primary buffers in blood are Hb in erythrocytes and bicarbonate in the plasma. This buffering by Hb is achieved by ionization of the imidazole ring of histidines in the protein.

Gas exchange

Optimal gas exchange requires a matched supply of inspired gas and blood to each alveolus; however, an inherent physiological ventilation (*V*)/perfusion (*Q*) mismatch exists in humans. Blood flow is greatest in the dependent bases when standing owing to hydrostatic pressure, whereas ventilation is reduced owing to orthostatic pressure. The converse applies at the lung apices. Respiratory diseases exacerbate the disparity between ventilation and perfusion, leading to respiratory failure. In chronic obstructive pulmonary disease (COPD), bronchial constriction impairs airflow and destruction of alveolar walls limits gas exchange. In lobar pneumonia inflamed oedematous alveoli inhibit diffusion and in pulmonary emboli (PE) sections of lung lack blood supply but remain adequately ventilated. In critical care, patients are sometimes positioned so that their healthy lung is in the dependent position to improve blood flow and thus gas exchange; an example of this is prone ventilation in acute respiratory distress syndrome (ARDS).

Mechanics of ventilation

During inspiration, thoracic cavity volume increases owing to the inferior movement of the diaphragm and elevation of the ribs by intercostal muscle contraction. In turn, intrapulmonary volume increases (owing to effective pleural adhesion of the thoracic wall and lung) and generates a relatively negative intrapulmonary pressure. The gradient between this and the atmospheric pressure draws air into the lungs. When performing thoracic or abdominal surgery in patients with neuromuscular disorders (poliomyelitis, myasthenia gravis, motor neurone disease) respiratory failure is predictable and can be reduced by protective measures such as head-up positioning, chest physiotherapy and nocturnal non-invasive ventilation. Similarly, chest wall and spinal deformity (kyphoscoliosis, pectus excavatum) may restrict ventilation and should be considered preoperatively.

During resting ventilation, expiration is passive. Elastic recoil causes compression of the alveoli, resulting in increased alveolar pressure and diffusion of gases outwards towards the lower atmospheric pressure. Forced expiration utilizes both thoracic and abdominal muscles.

Airflow within the airways is laminar during quiet breathing but becomes turbulent at high flow velocities; therefore, any state that increases the rate of breathing will increase the work of breathing. Airflow resistance is increased by narrowing of the airways, thus bronchoconstriction, secretions and mucosal oedema all restrict airflow.

Compliance is defined as the force required to overcome the combined resistance provided by the counteracting elastic properties of the chest wall, pleura, lungs and abdomen in order to alter lung volume. The relationship dictates that the lower the compliance, the greater the pressure required to achieve a given change in volume (‘stiff lungs’). Different diseases are associated with decreased or increased compliance (Table 9.1).

Table 9.1 Conditions associated with changes in lung compliance

Decreased compliance	Increased compliance
Pulmonary fibrosis	COPD
Asthma Pneumonia	The ageing lung
Atelectasis/lobar collapse	
Pulmonary oedema	
Obesity	
Chest wall deformity	
Thoracic surgery	

COPD, chronic obstructive pulmonary disease.

Control of ventilation

Respiratory rate and tidal volume are controlled to maintain alveolar tensions of O_2 in the region of 13.3 kPa and CO_2 at 5.5 kPa and a consequent arterial pH of 7.45. Central (located on the ventral surface of the medulla) and peripheral (located in the carotid and aortic bodies) chemoreceptors detect increases in CO_2 and H^+ concentration in cerebrospinal fluid and arterial blood, and decreases in arterial O_2 tension to stimulate hyperventilation via the brainstem respiratory centre (located in the pons and medulla). Local mechanical receptors such as stretch, J and irritant receptors also influence ventilation through feedback mechanisms. The cerebral cortex can exhibit voluntary control over breathing.

Innervation of the lungs

The pleura and the lungs are innervated by anterior and posterior pulmonary plexi, which consist of parasympathetic (vagal) and sympathetic fibres. Parasympathetic efferent fibres control smooth muscle bronchoconstriction, vasoconstriction and gland secretion. Conversely, sympathetic efferent fibres relax and inhibit these processes. Visceral sensory fibres sense touch, pain and stretch.

The diaphragm is innervated by spinal nerves C3–5 via the phrenic nerves. The intercostal muscles are supplied via the intercostal nerves from spinal nerves T1–11.

Phrenic nerve damage is a serious, well-recognized complication of thoracic procedures which results in diaphragm paralysis. The diagnosis is made with a chest radiograph (raised hemidiaphragm), thoracic ultrasound ('sniff test' demonstrates paradoxical diaphragm movement) and nerve conduction studies. Lung function testing can be useful (low volumes and worse in supine position). Management includes head-up positioning and non-invasive ventilation. Recovery is possible but is usually protracted given the slow rate of nerve regeneration.

Respiratory failure

Respiratory failure is defined as a PO_2 of less than 8 kPa or an inability of the lungs to adequately oxygenate arterial blood and/or clear CO_2.

There are four mechanisms of respiratory failure:

1. *shunting* (pneumonia, atelectasis, congenital heart disease)
2. *diffusion impairment* (pneumonia, interstitial lung disease)
3. *hypoventilation* (neuromuscular disease, drug overdose, chest wall compromise such as traumatic non-penetrating injury, diaphragm splinting, fatigue, wasting)
4. *ventilation–perfusion mismatch* (COPD, PE, pulmonary oedema, interstitial lung disease).

The type of respiratory failure is defined according to the arterial CO_2 concentration. *Type 1 respiratory failure* is hypoxia (PO_2 <8 kPa) in the presence of normocarbia. This can be considered to be a failure of oxygenation. This is often corrected with O_2 via a facemask but sometimes requires continuous positive airways pressure (CPAP) or invasive ventilation. Shunting and diffusion impairment are associated with type 1 respiratory failure. *Type 2 respiratory failure* is hypoxia (PO_2 <8 kPa) with associated hypercarbia (PCO_2 >6.5 kPa) and indicates a failure of ventilation. The presence of an acidosis suggests an acute deterioration. Metabolic compensation occurs in chronic type 2 respiratory failure, thus the pH is normalized but bicarbonate elevated. Although mechanical ventilation was previously the only option to manage such cases, bilevel non-invasive ventilation in which an inspiratory positive airways pressure inspiratory pressure (IPAP) is applied in addition to providing positive end expiratory pressure via a facial or nasal mask is now widely used, although the evidence base for its use in postoperative cases is limited. Non-invasive ventilation is indicated when there is type 2 respiratory failure in association with a pH <7.35.

Non-invasive ventilation

Non-invasive ventilation is generally applied in the setting of a high-dependency or intensive care unit, as this requires specialist staff and equipment. It is important to be aware that the raised intrathoracic pressures generated by non-invasive ventilation cause reduced venous return and thus cardiac output. Care is needed to ensure that the patient is intravascularly replete to avoid tissue hypoperfusion and hypotension, which in itself can worsen acidosis.

Non-invasive ventilation provides time to establish definitive treatment for the underlying condition and is not in itself a panacea for deterioration. Once non-invasive ventilation has been established it is essential to deduce the cause for the respiratory deterioration and instigate targeted medical therapy. Early recognition and treatment of atelectasis, bronchospasm, infection and hypoxia with nebulizers, antibiotics, chest physiotherapy and appropriate O_2 therapy can prevent the need for mechanical ventilation.

The contraindications to non-invasive ventilation include:

- reduced Glasgow Coma Scale score (<8)
- cardiac or respiratory arrest
- severe hypotension
- severe vomiting/haematemesis/excess bronchial secretions
- inability to protect own airway or preponderance to upper airway obstruction, including large head and neck tumours and anaphylaxis.

Preoperative assessment

Preoperative assessment of patients with respiratory disease aims to establish baseline function and optimize medication prior to surgery (Table 9.2). The patient's 'normal' exercise tolerance should be established. Establishment of predisposition to CO_2 retention is valuable preoperatively via arterial blood gas sampling. Useful information can be gained from the patient's response to previous surgical procedures and the time to recovery.

Table 9.2 Pulmonary function test patterns in respiratory disease

	Obstruction, e.g. COPD	**Intrapulmonary restriction, e.g. ILD**	**Extrapulmonary restriction, e.g. obesitypleural, chest wall and neuromuscular disorders**
FEV_1	↓↓	↓	↓
FVC	↓ or ↔	↓	↓
FEV_1/FVC ratio	↓	↑ or ↔	↑ or ↔
TLC	↑ or ↔	↓	↓
RV	↑	↓	↑
TLCO	↓	↓	↓ or ↔
KCO	↓	↓	↑ or ↔

↓, reduced, ↑, increased, ↔, normal. COPD, chronic obstructive pulmonary disease; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; ILD, interstitial lung disease; KCO, carbon monoxide gas transfer coefficient RV, residual volume; TLC, total lung capacity; TLCO, carbon monoxide gas transfer capacity.

Risk factors for postoperative respiratory complications include: underlying cardiorespiratory disease, older age, long anaesthetic time, smoking history, thoracic and abdominal surgery, obesity and poor functional baseline. Identification of risk factors and appropriate optimization with bronchodilators, antibiotics, steroids, pulmonary rehabilitation and non-invasive ventilation may improve postoperative recovery. Smoking cessation may reduce the risk of hospital-acquired infection and thus should be strongly encouraged preoperatively. Ideally this should occur at least 6 weeks prior to surgery.

Investigations in the preoperative assessment of cardiorespiratory patients include:

- O_2 saturation (if <95% on room air an O_2 assessment is required with arterial blood gases)
- chest radiograph
- spirometry [forced expiratory volume in 1 second (FEV_1), forced vital capacity (FVC), gas transfer] with reversibility studies
- 6 minute walk test
- consider cardiac comorbidity with ECG and echocardiogram
- cardiopulmonary exercise test.

Management of postoperative dyspnoea

Postoperative breathlessness is a common complication especially following thoracic and abdominal surgery. The main causes are pulmonary oedema, pneumonia, exacerbation of underlying lung disease and thromboembolic disease. The remainder of this chapter focuses on the recognition and management of respiratory disease in the context of the postoperative patient.

Investigations of postoperative dyspnoeic patient include:

- history and examination, including review of preoperative lung function tests
- observations of vital signs and O_2 saturation
- ECG
- blood tests: full blood count and C-reactive protein
- arterial blood gases
- chest radiograph
- consider CT thorax with/without CT pulmonary angiography (CTPA).

The common causes of postoperative dyspnoea are:

- Respiratory:
  - pneumonia
  - atelectasis
  - pleural effusion
  - pneumothorax
  - exacerbation of underlying lung disease (e.g. COPD)
  - PE.
- Cardiac:
  - left ventricular failure/pulmonary oedema
  - acute coronary syndrome.
- Others:
  - anaemia
  - iatrogenic phrenic nerve damage
  - fat embolism
  - anaesthetic/analgesic-related hypoventilation.

Pulmonary infection

Pneumonia is common, associated with significant morbidity and mortality and is responsible for prolonged hospital stay and cost. Those at increased risk include smokers, COPD patients, the elderly and immunosuppressed individuals.

Community-acquired pneumonia

This is most commonly caused by *Streptococcus pneumoniae*, *Haemophilus influenzae* or *Mycoplasma*. Treatment should be guided by local policy but is often with amoxicillin and clarithromycin for 7 days. Pre-existing community-acquired pneumonia (CAP) should be considered during admission clerking of all elective and emergency patients as it can complicate postoperative recovery if overlooked.

Hospital-acquired (nosocomial) pneumonia

This is defined as the presence of new radiographic infiltrates and sepsis occurring more than 48 hours following hospital admission and is associated with mortality rates between 20% and 50%. The causative pathogens include *Staphylococcus*, *Klebsiella* and *Pseudomonas aeruginosa* as well as those associated with CAP. Inadequate cough secondary to pain and muscular weakness leads to atelectasis, thus adequate analgesia, early mobilization and physiotherapy are important preventative measures. Treatment should be determined by local microbiology policies but is usually with broad-spectrum antibiotics providing anaerobic and Gram-negative cover such as piperacillin/tazobactam or second/third-generation cephalosporins.

Aspiration

Aspiration is a significant postoperative cause of nosocomial pneumonia. The risk of aspiration of oral secretions, food, fluid and medications is raised by impaired consciousness (sedatives, opiate analgesia, anaesthesia) and impaired protective cough/gag reflexes (postextubation, head, neck and facial surgery, nasogastric feeding). Aspiration should also be considered in trauma patients. Nasopharyngeal bleeding, loose teeth, alcohol/drug intoxication and brain injury should raise suspicion.

Clinical management

Clinical features include fever, cough, sputum, pleurisy, dyspnoea and confusion. Examination findings are tachycardia, tachypnoea, hypoxia and focal chest signs (bronchial breathing, crepitations, reduced air entry, increased vocal fremitus and decreased percussion note). Chest radiograph, sputum and blood cultures, arterial blood gas and blood tests (C-reactive protein, full blood count, renal and liver function tests) are necessary to guide diagnosis and management. Treatment with antibiotics should be guided by local policy and initiated immediately following the collection of microbiological samples. Antibiotic therapy can subsequently be rationalized when culture results become available. Adequate oxygenation ($PO_2 \geq 8$ kPa) and hydration should be maintained. Consideration should be given to chest physiotherapy, deep breathing and saline nebulizers to facilitate effective secretion clearance. Arterial blood gas sampling should be used to assess treatment response and severity. Failure to oxygenate above a PO_2 of 8 kPa with high-flow O_2 and the presence of a metabolic acidosis indicate severe disease and the need for escalation of care. Early identification of complications including shock, parapneumonic effusion, lobar collapse, abscess, empyema and ARDS is important and should prompt referral to respiratory or intensive care physicians.

Acute respiratory distress syndrome

ARDS is the development of acute lung inflammation and increased vascular permeability secondary to proinflammatory mediator release triggered by a variety of insults including sepsis, trauma, smoke inhalation, fat emboli, blood transfusion and pancreatitis. Mortality is greater than 50% and lung fibrosis common among survivors. Diagnosis requires the presence of sudden onset diffuse bilateral infiltrates on chest radiograph, a PaO_2/FiO_2 ratio of less than 200 mmHg and the absence of left ventricular failure (pulmonary wedge artery pressure <18 mmHg). Treatment of the underlying cause and supportive management are the mainstay. Mechanical ventilation is almost always required and evidence supports the use of low tidal volumes between 6 and 8 mL/kg and plateau pressures of <30 cmH_2O to minimize potential barotrauma. Other therapies that may be of benefit but that are not proven to be efficacious include inhaled nitric oxide (pulmonary vascular smooth muscle relaxant), methylprednisolone, fluid restriction, prone ventilation and extracorporeal membrane oxygenation (ECMO).

Exacerbation of airway disease: asthma/chronic obstructive pulmonary disease

Clinical management

Pre-existing airway obstruction may worsen peri- and postoperatively. It is important to optimize these patients preoperatively when there is an opportunity to do so in conjunction with respiratory physicians. Ensure that regular medication, and especially inhaled therapy such as inhaled corticosteroids, is continued throughout. Individuals with severe COPD (FEV_1 <30% or carbon monoxide gas transfer capacity <30% or on long-term O_2 therapy) are particularly at risk and need careful monitoring and also consideration of postoperative intensive or high-dependency care.

The use of nebulized β-agonists (such as salbutamol) and anticholinergics (such as ipratropium) needs to be considered if there is an exacerbation of underlying disease. In some cases it may be necessary to commence corticosteroids in addition to treating any infective component. Type 2 respiratory failure can occur especially in patients with pre-existing and severe COPD/emphysema.

Pulmonary embolism

Virchow's triad (venous stasis, vascular endothelial injury, hypercoagulability) explains the high incidence and mortality of venous thromboembolism (VTE) among postoperative patients. Most general surgical patients possess multiple risk factors for developing PE (Figure 9.2), such as major abdominal/pelvic surgery, immobility, dehydration, critical care admission, central venous access, malignancy, previous VTE, obesity, medical comorbidity and older age.

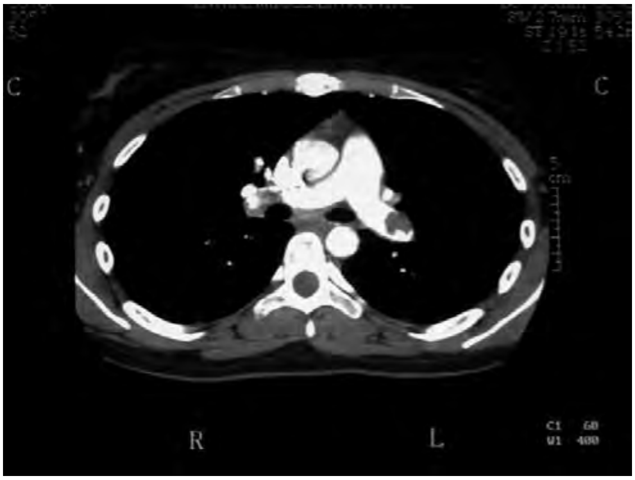


Figure 9.2 Acute pulmonary emboli in left main pulmonary artery.

Mechanical and pharmacological VTE prophylaxis are recommended for emergency admissions with inflammatory conditions and all surgical patients undergoing a procedure with an anaesthetic time of >90 minutes (>60 minutes if involving lower limbs or pelvis) or predicted prolonged immobility. Bleeding risk and contraindications to prophylaxis must be assessed on an individual basis. National guidance also suggests consideration of discontinuation of oral contraception or hormone replacement therapy 1 month preoperatively.

PE should be considered in patients with dyspnoea, pleuritic pain, haemoptysis, haemodynamic instability or collapse. Examination may be normal but tachycardia, new onset atrial fibrillation (AF) and tachypnoea are common. Other features include hypoxia, low-grade fever, hypotension, pleural rub, right heart failure and the presence of deep vein thrombosis. If there is high clinical suspicion and no contraindications, patients should receive treatment dose anticoagulant therapy immediately. An ECG may show sinus tachycardia, AF or right bundle branch block. The '$S_1Q_3T_3$' pattern is rare. Routine blood tests are helpful to exclude alternative causes; however, a D-dimer test is of no value in the surgical patient as this will be elevated postoperatively. An arterial blood gas often shows an increased A–a gradient and respiratory alkalosis (in response to the hypoxia, hyperventilation causes hypocarbia but is unable to increase oxygenation owing to the ventilation/perfusion mismatch) but can be normal. A chest radiograph may reveal an alternative pathology. Pleural effusions and wedge infarction may be seen with PE. A CTPA is diagnostic but a lower limb Doppler ultrasound can be useful in patients in whom a CTPA is contraindicated (contrast allergy, renal failure), as 70% of patients with PE will have an associated lower limb thrombus.

The management of PE is with O_2 and anticoagulant therapy. Low-molecular-weight or unfractionated heparin have equal efficacy. The total duration of anticoagulation following postoperative VTE is 4–6 weeks unless there is an underlying chronic risk factor (e.g. malignancy, previous idiopathic VTE, thrombophilia) and is usually with warfarin. In massive PE with collapse, thrombolysis should be considered. It is sometimes difficult to obtain a confirmatory CTPA in these circumstances and hence assessment of possible right ventricular strain or right-sided volume overload with echocardiography may be possible and in itself may identify intracardiac or intrapulmonary thrombus. Embolectomy is an alternative if life-threatening haemorrhage is likely as a result of anticoagulation. Inferior vena caval filters have not been proven to prevent fatal PE and these should not be used routinely.

Pleural effusion

Pleural effusions (Figure 9.3) are a common postoperative finding that occur for a wide variety of reasons (Table 9.3). Presentation of these can be gradual or sudden and is with dyspnoea, pleuritic chest pain or referred pain to the shoulder tip. Signs include reduced air entry, decreased percussion note, decreased vocal resonance/tactile fremitus, reduced expansion and bronchial breathing above the fluid level.

It is vital to establish the underlying cause of the effusion. All new pleural effusions should be sampled via ultrasound-guided thoracocentesis and sent for analysis of protein, albumin, lactate dehydrogenase, glucose and pH levels as well as microscopy, culture and cytology. Specialist tests include pleural fluid amylase and triglycerides, which can be raised in pancreatitis-associated effusions, oesophageal rupture and chylous effusions. The exception is when there is substantial risk of introducing infection such as after oesophagectomy or when there is overlying venous or skin grafting to the posterior thorax. In such cases pleural instrumentation should be avoided unless clinically indicated. Effusions are either transudative or exudative (Table 9.4) and the causes of each are given in Table 9.3.

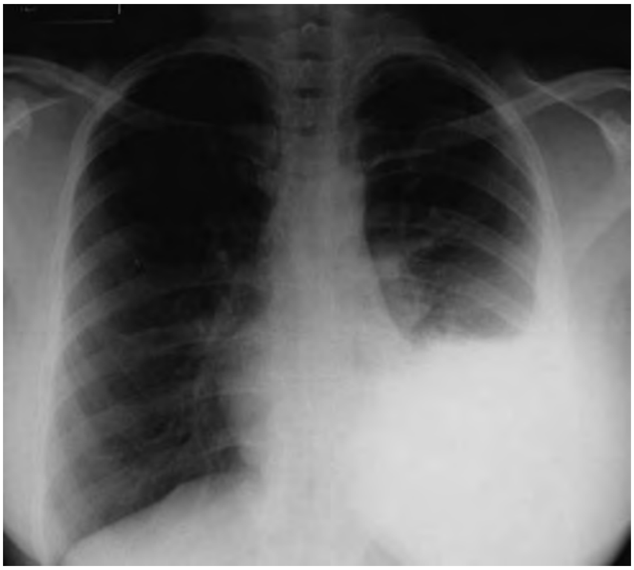


Figure 9.3 Left pleural effusion.

Table 9.3 Causes of pleural effusions

Causes of transudative effusion	Causes of exudative effusion
Increased venous pressure Left ventricular failure Constrictive pericarditis Fluid overload	Increased capillary permeability Infection (pneumonia, empyema, subphrenic collection, TB, oesophageal rupture) Inflammation (PE, rheumatoid arthritis, systemic lupus erythematosus, pancreatitis) Malignancy
Decreased oncotic pressure Hypoalbuminaemia (malnutrition, critical illness, malabsorption) Renal failure (nephrotic syndrome) Hepatic failure	Impaired lymphatic drainage Chylothorax (iatrogenic damage to thoracic duct, lymph invasion secondary to tumour)

PE, pulmonary embolism; TB, tuberculosis.

Table 9.4 Transudative versus exudative pleural effusions

	Transudates	Exudates
Pleural fluid protein/serum ratio	<0.5	>0.5
Pleural fluid LDH/serum ratio	<0.6	>0.6
Pleural fluid LDH level in relation to upper limit of normal for serum	<2/3	>2/3

LDH, lactate dehydrogenase.

Transudates infrequently require drainage as reaccumulation occurs too rapidly to be of benefit. In these cases treatment should focus on correction of the underlying condition, such as augmentation of nutrition in hypoalbuminaemia or treatment of cardiac failure/fluid overload. Exudates also require medical management of the underlying pathology but often require concurrent intercostal drainage and it is especially important to ensure that an empyema has not occurred (suspicion should be raised if there is evidence of ongoing sepsis), as this definitely requires drainage as antimicrobials are unable to penetrate and hence resolve this area of sepsis. Indicators of an empyema are aspirates of frank pustular material, presence of organisms on microscopy, polymorphonuclear cells and a pH <7.1.

Pneumothorax

Pneumothoraces (air within the pleural space) (Figure 9.4) can be either primary (occurring in normal lung) or secondary (underlying lung disease) and may occur spontaneously or following trauma. There is a strong association with smoking, males and small airways disease.

The classic presentation is with sudden onset breathlessness or pleuritic chest pain. Signs include tachycardia, tachypnoea, hypoxia and reduced expansion and breath sounds on the affected side. However, these features are frequently absent, especially in smaller pneumothoraces. A chest radiograph is usually diagnostic, but a CT thorax (Figure 9.5) is sometimes required to detect loculated pneumothoraces, to differentiate bullous lung disease from a pneumothorax and in trauma situations.

A 2 cm rim of air equates to a 50% lung collapse and it is at this volume that aspiration or drainage is required. In patients younger than 50 years with no underlying lung disease, aspiration should be attempted up to two times before drainage is considered. In patients with COPD, aspiration is rarely successful, thus drainage should be first-line management. Small bore Seldinger-type drains are usually sufficient.

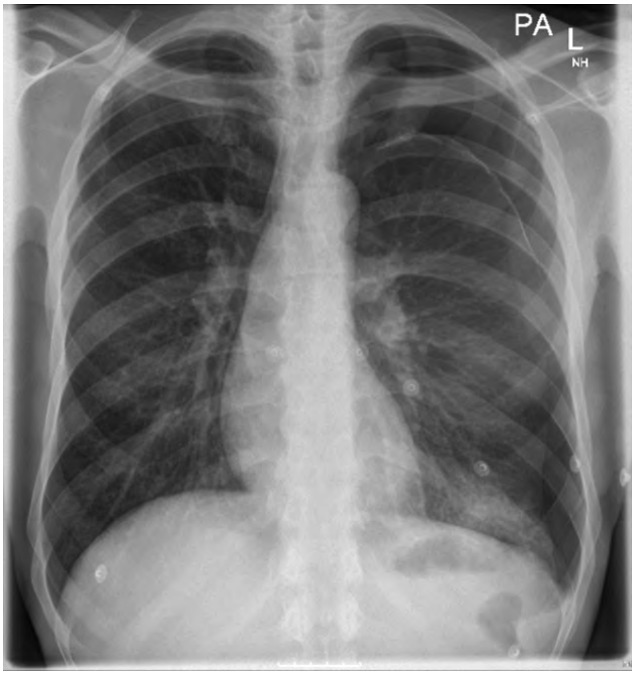


Figure 9.4 Left-sided pneumothorax.

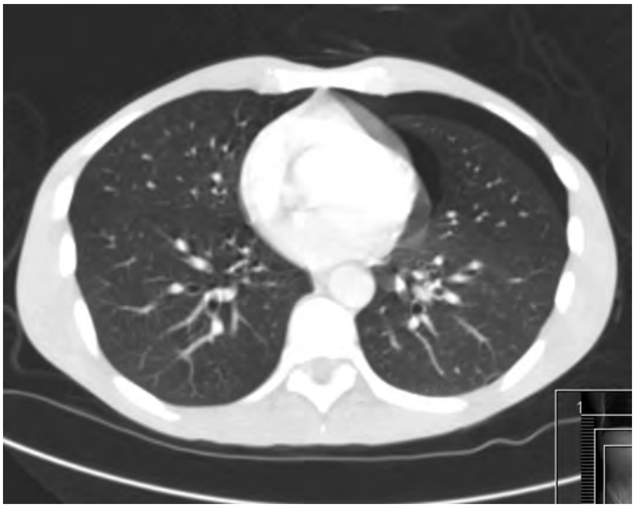


Figure 9.5 CT scan of left pneumothorax.

Traumatic pneumothoraces invariably require intercostal drainage and large bore Argyll-type drains are necessary. Small pneumothoraces (<2 cm rim of air) can be managed with observation alone. High-flow O_2 (unless contraindicated) is used to increase the rate of resorption of the pneumothorax and thus should be prescribed for all patients with pneumothorax. Advise patients that diving is contraindicated without a definitive surgical procedure and that air travel should be avoided for 7 days following confirmation of resolution with chest radiograph (the British Thoracic Society recommends a period of 2 weeks following a traumatic pneumothorax).

Haemothorax

Blood within the pleural cavity should be suspected if there is a pleural collection, especially following a thoracotomy or trauma and the patient becomes breathless, hypotensive and anaemic. As above, this should be aspirated diagnostically under ultrasound guidance. If a haemothorax is confirmed it is necessary to insert a large bore intercostal drain (minimum 28 G) with an underwater seal bottle to improve ventilation. Transfusion may also be required to correct the resulting anaemia and fluid resuscitation is important. In cases in which there is insufficient clearance, a significant clotted haemothorax or inability to re-expand the lungs, refer for a cardiothoracic review and decortication.

Re-expansion pulmonary oedema

This rare condition can occur following rapid re-expansion of a collapsed lung by drainage of pleural effusion or pneumothorax. It is associated with a significant mortality of up to 20%. Dyspnoea, increasing chest pain, hypoxia and the development of crepitations following intercostal drain insertion suggests the diagnosis. A chest radiograph usually demonstrates ipsilateral pulmonary infiltrates; however, re-expansion pulmonary oedema can occur bilaterally or contralaterally. Treatment is supportive with high-flow O_2 and CPAP or mechanical ventilation. Diuretics are contraindicated as this may worsen the condition. Risk factors include younger age, total lung collapse, mature pleural collections and rapid drainage. Controlled drainage of effusions (aim for maximum 1 litre per day) and avoidance of negative pressure suction in the first 48 hours of drainage reduces the risk.

GUIDE TO FURTHER READING

Bernard GR, Artigas A, Brigham KL, *et al.* The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994;**149**(3 Pt 1):818–824.

British Thoracic Society. Pleural disease guideline 2010. *Thorax* 2010;**65**(suppl. II):ii4–ii76.

Lim WS, Baudouin SV, Hill AT, *et al.* BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax* 2009;**64**:1–55.

National Institute for Health and Clinical Excellence. *Reducing the Risk of Venous Thromboembolism (Deep Vein Thrombosis and Pulmonary Embolism) in Patients Admitted to Hospital.* Clinical guideline 92. London, UK: NICE, 2010. Available from: www.nice.org.uk/guidance/CG92.

CHAPTER 10

Renal insufficiency

NEILL DUNCAN AND RICHARD W. CORBETT

Introduction

The prime function of the kidney is to assist with the maintenance of the internal environment through the control of four broad physiological functions: (1) the selective control of the elimination or retention of water, (2) the selective control of the elimination or retention of electrolytes and other solutes, (3) maintenance of pH homeostasis and (4) hormone production. The first three of these functions occur through the combination of glomerular ultrafiltration, tubular reabsorption and active secretion along the length of the nephron, leading to the excretion of urine, whereas a number of hormonal processes, most notably erythropoietin (EPO) production and the activation of vitamin D, occur away from the glomerular and tubular apparatus.

Renal insufficiency relates to the inability of the kidneys to perform these functions in either the acute short term or the chronic long term, and damage may be reversible or permanent. The label of acute renal failure has been dropped in favour of acute kidney injury (AKI) as this terminology allows for states in which the kidney has not failed but may be at risk or injured, and in which early recognition and correct management may result in renal function being recovered. AKI, even at its earliest stages, carries an attributable risk of mortality. The physiological consequences of AKI can be managed by medical means in the earlier stages; it is only when organ failure has occurred that renal replacement therapy (RRT), for example haemodialysis or haemofiltration, is frequently required as a life-sustaining therapy. Although these treatments are effective and can approximate the basic physiological functions of the kidney, they are not without problems and associated morbidity; it is far more desirable to avoid progressive significant renal injury and to maintain native renal function.

In the surgical patient it is likely that renal insufficiency is encountered either as a pre-existing comorbidity as chronic kidney disease (CKD) or if AKI develops, because of either the surgical disease or an intervention. It is clearly possible for both of these states to coexist; indeed, a patient with pre-existing CKD is at greater risk of AKI.

This chapter will deal with AKI and CKD in turn but with a particular emphasis on AKI given its particular relevance to surgical disease and the critically unwell patient. Of particular clinical importance is the need to liaise closely with a nephrologist at an early stage in the care of patients who are receiving RRT with either chronic haemodialysis or renal transplantation. These two groups have particularly complex and differing needs, which are outside the scope of this chapter, but if they are not attended to diligently they can result in significant complications.

Acute kidney injury

Definition and epidemiology

AKI is characterized by a rapid deterioration in renal excretory function resulting from a wide range of causes. This is evidenced by oliguria and a rise in serum markers of excretory function, such as creatinine, over a period of hours to days. An important caveat is the time lag between the onset of renal injury and the rise in serum creatinine, which can be up to 36 hours, invalidating the use of calculations that estimate glomerular filtration rate (GFR) in this setting. The reciprocal correlation of serum creatinine with GFR, moreover, means that, by the time a rise in creatinine is seen to the upper limit of 'normal', up to 50% of renal function may have been lost.

Multiple definitions of AKI exist, although, in practice, these are not used to guide therapy but are descriptive of the risk of associated complications. The Acute Kidney Injury Network (AKIN) defines (Table 10.1) the stage of AKI either as an absolute increase or proportional increase in creatinine from baseline or as a change in urine output; for example, AKI stage 1 is defined as an

Table 10.1 The Acute Kidney Injury Network (AKIN) criteria for acute kidney injury

Stage	Serum creatinine criteria	Urine output criteria
1	Increase in serum creatinine of ≥0.3 mg/dL or increase to ≥150–200% from baseline	<0.5 mL/kg/h for >6 hours
2	Increase in serum creatinine to >200–300% from baseline	<0.5 mL/kg/h for >12 hours
3	Increase in serum creatinine to >300% from baseline or serum creatinine of ≥4 mg/dL; or receiving renal replacement therapy	<0.3 mL/kg/h for 24 hours or anuria for 12 hours

Table 10.2 The Risk, Injury, Failure, Loss and End-stage Kidney (RIFLE) criteria for acute kidney injury

Stage	GFR criteria	Urine output criteria
Risk	Increased serum creatinine ×1.5 baseline or GFR decrease >25%	<0.5 mL/kg/h for >6 hours
Injury	Increased serum creatinine ×2 baseline or GFR decrease >50%	<0.5 mL/kg/h for >12 hours
Failure	Increased serum creatinine ×3 baseline or GFR decrease >75% or serum creatinine ≥4 mg/dL with an acute increase of ≥0.5 mg/dL	<0.3 mL/kg/h for 24 hours or anuria 12 hours
Loss	Persistent acute kidney injury, i.e. complete loss of kidney function >4 weeks (i.e. dialysis dependence for 4 weeks)	
End-stage	End-stage kidney disease (i.e. dialysis dependent for >3 months)	

GFR, glomerular filtration rate. Note: 4 mg/dL is equivalent to 354 μmol/L; 0.5 mg/dL is equivalent to 44 μmol/L; 0.3 mg/dL is equivalent to 26 μmol/L.

increase in serum creatinine of >30 mmol/L, a 150–200% increase in creatinine from baseline or >6 hours of oliguria (<0.5 mL/kg/h) in patients who are adequately hydrated. The Risk, Injury, Failure, Loss and End-stage Kidney (RIFLE) classification criteria (Table 10.2) are based on patients admitted to an intensive care unit. The acronym mixes three grades of increasing severity of AKI based on similar criteria to those used in the AKIN definition (risk, injury, failure) but also includes two outcome variables (loss and end-stage renal disease). The entry criteria in both of these definitions may appear to be very low levels of renal impairment, but are strongly associated with evidence of poor outcomes and should act as a trigger for urgent investigation and corrective intervention.

Although some forms of surgery will entail a higher risk, AKI occurs in around 1% of all non-cardiac surgery patients who had normal preoperative renal function, with 0.1% requiring RRT. This accounts for between 20% and 50% of all hospital-associated AKI. Even in those patients with early stages of AKI there is an association with increased mortality, length of stay and use of resources in hospitalized patients. In cohort studies, stage 1 AKI is associated with a two- to fourfold increase in all hospital mortality. AKI not only occurs in association with failure of other organs but emerging evidence suggests that it also leads to dysfunction of other organs, possibly through upregulation of inflammatory mediators and endothelial injury.

The impact of renal impairment is not restricted to hospital admission; patients with initially normal renal function who recover after postoperative AKI will have an increased risk of death of 20% in the longer term compared with patients without AKI, as well as an increased risk of developing CKD. As might be expected, in those with pre-existing CKD the risks are greater, with a fivefold increased risk of prematurely requiring long-term RRT.

Prevention

In-hospital AKI arises frequently from a combination of insults: haemodynamic instability, loss of renal autoregulation, an inflammatory state and direct nephrotoxicity. Identifying the patient at risk and attempting to modify the factors that may lead to AKI are recommended given the evidence that adverse outcomes are associated with even minor levels of renal impairment. A variety of clinical features are known to be highly associated with an increased risk of AKI (Table 10.3); although it is possible to curtail the risk of AKI through the modification of some of these risk factors, many are unmodifiable and are associated with altered physiology leading to greater haemodynamic instability and impaired autoregulation.

Table 10.3 Features associated with increased risk of acute kidney injury

General	Perioperative
Age >60	Emergency surgery
Male	Intraperitoneal surgery
Congestive cardiac failure	Aortic cross-clamping
Liver cirrhosis	Massive blood transfusion
Hypertension	
Chronic kidney disease	
Diabetes mellitus (types 1 and 2)	
Active infection	

A clinical assessment of AKI risk should be made in all patients admitted to hospital as an emergency as well as in those undergoing significant elective procedures. Those identified as high risk should receive more intensive clinical and biochemical monitoring. Preoperative identification of high-risk patients, where possible, allows time for investigation of previously unidentified CKD, discontinuation of nephrotoxic agents and liaison with an anaesthetist and nephrologist. This is paramount for those with advanced renal impairment in whom significant interventions may precipitate the need for long-term RRT.

Minimizing haemodynamic instability

Kidneys receive around 20% of all cardiac output and they are particularly vulnerable to small reductions in blood flow. Maintaining renal perfusion is central to the prevention of AKI and most of the risk factors for AKI cause damage by potentiating hypoperfusion. An adequate blood pressure, although an element of treatment, is not the main goal of treatment by any means. The key to management is maintaining blood flow through the kidney by maintaining cardiac output with an appropriate circulating volume and vessel patency to the viscera.

Optimizing the circulating volume is difficult in AKI with surgery or trauma as there is reduced ability to excrete both sodium and water and increased risk of iatrogenic fluid and salt loading with the injudicious use of sodium-rich fluid with prescription of fluid replacement or drugs such as antibiotics or parenteral nutrition. The use of sodium-deplete fluids may lead to hyponatraemia. The tissue oedema that arises from aggressive use of any intravenous fluid is associated with poor wound healing, reduced mobility, venous stasis and thrombosis, pulmonary oedema and infection, and, ultimately, poorer outcome.

The choice and use of intravenous fluids remain a controversial area for the surgical patient but have recently been reviewed and summarized in UK consensus guidelines (British Consensus Guidelines on Intravenous Fluid Therapy for Adult Surgical Patients). These make a large number of recommendations, most notably encouraging in general the use of 'physiological' resuscitative or maintenance fluids such as Ringer's lactate/acetate or Hartmann's solution to avoid the large sodium and chloride loads associated with 0.9% 'normal' saline. There is some evidence that the use of hydroxyethyl starch (HES) preparations as colloid volume expanders may increase the risk of AKI in susceptible populations through direct tubular toxicity and are best avoided in this group of patients.

Those patients who are at greatest risk of hypovolaemia, such as those with significant intra-abdominal fluid losses (e.g. pancreatitis, bowel obstruction), extensive burns or kept nil by mouth for extended periods of time, need closer attention to their fluid requirements to ensure an adequate circulating volume. Patients with significant sepsis or cardiac dysfunction may require high-dependency care in order to maintain adequate renal and other organ perfusion through the use of vasopressors and inotropes.

There is no evidence that the use of 'low-dose' dopamine or fenoldopam (to preserve renal perfusion), mannitol (to force an osmotic diuresis) or the use of sodium bicarbonate fluid replacement have any benefit in preventing AKI.

Nephrotoxins

A wide variety of medications have nephrotoxic effects (Table 10.4) and are best avoided when possible or steps taken to limit the exposure when their use is unavoidable. Ideally they should be omitted 24–48 hours prior to surgery or other procedure (including contrast-mediated examinations) and recommenced once the risk of renal injury has passed.

Non-steroidal anti-inflammatory drugs (NSAIDs) can precipitate renal injury in a number of ways but predominantly do so by impairing renal autoregulation through the inhibition of cyclo-oxygenase, which mediates prostaglandin-induced vasodilatation of the afferent glomerular arteriole. Their beneficial analgesic effects need to be balanced against the risk of AKI; they are contraindicated in the high-risk population. Angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers are widely used in the management of CKD as they have strong prognostic benefits in the long term by reducing glomerular hypertension. However, in the acute setting their mechanism of action causing efferent arteriolar dilatation will exaggerate any haemodynamic insult by reducing the ability of the kidneys to autoregulate blood flow

Other hypertensive medications and diuretics are not in themselves nephrotoxic and may not need to be discontinued. In an acutely sick patient prone to hypotension their use needs to be reviewed on a frequent basis to reduce the risk of shock. Perioperatively advice should be sought from an anaesthetist. Metformin, equally, is not nephrotoxic, but can accumulate in renal impairment and is often stopped in a prophylactic manner to avoid an ensuing type B lactic acidosis should AKI occur.

Limiting contrast nephropathy

Contrast-mediated radiological examinations and interventional procedures are increasingly common, as are their complications. Ultimately, clinically mandated examinations or procedures should not be deferred but instead a careful assessment of risk made with preventative measures instituted to limit the kidney injury that arises from their use.

The exact mechanism of contrast-induced nephropathy is unclear but may relate to a combination of renal vasoconstriction and direct nephrotoxicity of the contrast agent. Conventional contrast agents are based upon a tri-iodinated ionized benzene ring and have a high osmolality (approximately five times greater than plasma), which along with their viscosity may contribute to the toxicity of these agents. Newer iso-osmolar non-ionic agents such as iodixanol appear to be less nephrotoxic but are considerably more expensive. The use of these newer agents in those at risk of AKI, along with the minimization of the volume required, is preferable.

The other risk factors for contrast-induced nephropathy are similar to those for the development of AKI and in general the actions taken to prevent all forms of AKI are also true for contrast examinations, including ensuring euvolaemia and discontinuing nephrotoxic drugs as well as metformin 48 hours prior to contrast exposure. Specifically evidence suggests that the use of approximately 1–2 litres of either isotonic sodium chloride or sodium bicarbonate solutions starting a few hours prior to the use of contrast and continuing for several hours afterwards is beneficial, although the studies are diverse in the timing and volume used. However, the evidence for other preventative measures such as *N*-acetylcysteine is conflicting and limited, and this is not generally advocated.

Table 10.4 Some examples of common nephrotoxic medications

Class of medication	Examples	Mechanism of action
Anaesthetic agent	Enflurane, methoxyflurane	Tubular toxicity
Angiotensin-converting enzyme inhibitor	Ramipril, perindopril	Afferent arteriolar vasoconstriction
Angiotensin II receptor blocker	Irbesartan, losartan, candesartan	Afferent arteriolar vasoconstriction
Antifungals	Amphotericin	Distal tubular toxicity
Aminoglycosides	Gentamicin, amikacin	Accumulate in proximal tubules
Antivirals	Aciclovir, foscarnet, indinavir	Intratubular crystal formation and toxicity
Chemotherapeutic	Cisplatin, ifosfamide	Tubular cell toxicity
Immunosuppressant	Ciclosporin, tacrolimus	Intrarenal vasoconstriction
Non-steroidal anti-inflammatory	Diclofenac, ibuprofen, naproxen	Vasoconstriction and acute interstitial nephritis
Radiocontrast media	Iopanoic acid, iothalamate	Multiple sites of action

Renovascular disease

As well as causing hypertension, renal artery stenosis can be associated with an ischaemic nephropathy. The most common cause for this is an atherosclerotic process that may involve either or both renal arteries. Perioperative hypotension may exacerbate renal hypoperfusion in those patients in whom the stenoses are flow limiting, leading to a profound AKI. There is limited and controversial evidence for angiographic treatment of renal artery stenosis with balloon angioplasty and stenting when significant surgery is contemplated – particularly major cardiac or vascular procedures. The possible benefits of prophylactic intervention should be discussed with a nephrologist. If the need for surgery is immediate then this should not be postponed in order to treat renal artery stenosis.

Investigating the patient with acute kidney injury

In a patient with surgical disease in whom AKI is recognized early, it is often an early indication of systemic physiological disturbance. The concern is that these patients may deteriorate further and that the AKI is associated with other organ dysfunction. All patients should have prompt assessment by an experienced clinician, as appropriate intervention with regular reassessment may limit further complications; in those with multiorgan dysfunction this may be best delivered in a critical care environment.

In the first stages of care it may not be clear whether an abnormal serum creatinine represents an AKI or relates to long-standing CKD. Ideally, corroborative biochemical results from either primary care or previous admissions can clarify this but if these are not available then 12 hourly serial creatinine monitoring should be performed along with measurement of urine output. Although small renal size, anaemia and hypocalcaemia are all associated with untreated CKD, they can be misleading; for example, diabetes, myeloma, amyloid and infiltrative lymphoproliferative disease may be causes of CKD with preserved renal size. If there is any doubt, it should be assumed that the renal impairment relates to reversible AKI.

Attempting to establish the aetiology of the injury is important as it allows perpetuating factors to be treated. Traditionally, the causes of AKI have been divided into prerenal, intrinsic and postrenal disease, although frequently more than one insult has led to AKI. This approach also lends itself to concurrent investigation and treatment.

Prerenal

A fall in renal perfusion leading to a decrease in GFR is a common cause of AKI. The kidneys can compensate for modest reductions in perfusion by autoregulation to maintain glomerular filtration through a combination of afferent arteriolar vasodilation and efferent arteriolar vasoconstriction. As discussed above, NSAIDs and drugs that block the angiotensin system directly inhibit the afferent and efferent vascular homeostatic mechanisms, respectively, and therefore will reduce GFR. More severe reductions in renal perfusion overwhelm autoregulation and lead to a rapid deterioration in GFR. If identified and treated promptly, even drastic falls in renal perfusion are reversible; however, delay in resuscitation may lead to intrarenal injury with ischaemic acute tubular necrosis (ATN).

A fall in renal perfusion may occur with any of the causes of systemic shock: hypovolaemia, cardiac failure and inappropriate vasodilatation (e.g. sepsis and anaphylaxis), although the relatively sensitive nature of the kidneys to physiological insults means that AKI may be evident before circulatory collapse ensues. Some common causes for all of these are listed in Box 10.1. Cardiac failure either may occur as a pre-existing problem (congestive cardiac failure) or may be acute. It may be difficult to identify patients with cardiac failure but it is very important as aggressive expansion of the intravascular compartment is associated with a paradoxical reduction in cardiac output owing to alterations in the intrinsic cardiac inotropy in such patients. The result of falling cardiac output is a further fall in renal perfusion and pulmonary oedema. Both sepsis and liver cirrhosis lead to inappropriate vasodilation of the peripheries, compromising renal perfusion. Cirrhosis also causes significant splanchnic vasodilation – a mechanism that is implicated in the hepatorenal syndrome. Hepatorenal syndrome is an uncommon form of AKI in those with advanced liver disease that requires careful consultation with a liver centre as the only definitive treatment is liver transplantation.

Clinical assessment, including a thorough history and examination, is vital in the assessment of the patient with AKI. The key management questions at this stage are whether there is adequate renal perfusion and is there intravascular hypovolaemia? Oliguria is not necessarily evidence of hypovolaemia and in the short term in the postoperative patient may represent a normal physiological response. Clinical signs reflecting intravascular hypovolaemia include reduced capillary refill, poor peripheral perfusion, low jugular (central) venous pressure, and a falling trend in blood pressure and rising trend in pulse. However, assessing the postoperative patient or those patients with significant comorbidities can be complex; for example, a patient with cor pulmonale will normally have high right heart filling pressures and a jugular venous pressure that is visible may occur

BOX 10.1 Causes of prerenal acute kidney injury

- Extracellular fluid volume depletion
  - Increased gastrointestinal losses (vomiting, diarrhoea)
  - Increased renal losses (diuretics, diabetes insipidus)
  - Sequestration (rhabdomyolysis, pancreatitis, ileus)
  - Increased dermal losses (burns, excessive sweating)
- Intravascular volume depletion
  - Haemorrhage
- Decreased cardiac output
  - Cardiac failure (longstanding congestive cardiac failure, recent myocardial infarction)
  - Cardiac tamponade (large pericardial effusion)
- Systemic arterial vasodilation
  - Sepsis
  - Liver cirrhosis
- Renal artery stenosis

in the setting of intravascular hypovolaemia. More invasive monitoring with the use of central venous catheters and urinary catheters may be useful but are not mandated, and may carry risk, particularly of infection; their use should be carefully considered before insertion and de-escalation with timely removal is vital.

There is no evidence for the use of any particular crystalloid or colloid in the resuscitation of the patient with AKI. However, Hartmann's solution, which contains potassium, should be avoided in the presence of hyperkalaemia, and potassium supplementation when serum potassium is in the normal range should be routinely avoided to avoid iatrogenically creating a clinical need for dialysis in a patient with progressive AKI. Solutions containing high-molecular-weight HES are contraindicated in AKI. Once the circulating volume has been corrected, if there is persisting evidence of impaired systemic perfusion then the use of vasopressors or other adjuncts to maintain renal and central organ perfusion should be considered early. In the anuric patient, repeated challenges of parenteral fluids run the risk of life-threatening pulmonary oedema.

Postrenal

Renal obstruction can occur at any level from the renal pelvis, through the ureters and bladder to the urethra and may result from either extrinsic or intraluminal factors. The former is often due to a compressive tumour, retroperitoneal fibrosis or occasionally inadvertent ureteric ligation (either bilateral or unilateral). Intraluminal obstruction usually results from blood clots, calculi or bladder outflow disease. The resulting intratubular pressure within the kidney opposes glomerular filtration pressure and reduces renal blood flow leading to a fall in GFR.

In any patient with AKI, the possibility of obstruction should be considered early as prompt relief of the obstruction can result in a rapid amelioration of renal function and carries a good prognosis. As such, radiological imaging, preferably ultrasonography, should be performed within 24 hours of the diagnosis of AKI in all patients; ultrasound has 90–95% sensitivity for obstruction. The ultrasound may be falsely reassuring, however, and show non-dilated pelvicalyceal systems and ureters in spite of obstruction in certain conditions, i.e. patients who may have infiltrative or inflammatory disease of the ureters and patients who have established oilgoanuria in whom insufficient urine is generated to dilate the renal tract. A high index of suspicion should lead to attempts to disobstruct with percutaneous nephrostomy tube insertion in such cases.

Importantly, when clinically assessing a patient with AKI it should be remembered that obstruction does not necessarily lead to anuria; acute unilateral ureteric obstruction may result in a rising serum creatinine without a significant change in urine output, hence the need to seek radiological imaging in all patients with AKI. The technique used for the relief of an acute obstruction will depend upon the level and nature of the obstruction, but timely disobstruction should be sought to prevent the irreversible kidney failure that results from delay. Disobstruction is mandatory if there is any suggestion of infection upstream of the obstruction with pyonephrosis.

Intrinsic renal injury

This is practically considered last as a diagnosis of exclusion made once the patient is euvolaemic and in the absence of obstruction. It includes a great diversity of different pathologies from the most common cause of AKI, i.e. ATN, through the range of renal-specific and systemic diseases that cause glomerular and tubulointerstitial pathology, and that are primarily the remit and interest of the nephrologist. Although a presumptive diagnosis of ATN is often made on clinical grounds, it is a histopathological diagnosis and what is described as ATN in a hospitalized patient may actually represent other intrinsic renal disease. In addition to ATN, this section will focus on a handful of other common specific presentations of AKI that are commonly associated with surgical disease.

Acute tubular necrosis

ATN is rarely a result of a prolonged hypoperfusion injury alone but in the unwell patient arises from a combination of ischaemia, sepsis and nephrotoxic insults. Cells within the proximal tubule and thick ascending limb of the kidney, which lie within the medulla of the kidney, appear most susceptible to injury as a result of a number of factors. First, the majority of renal blood flow is directed towards the cortex, wherein lie the glomeruli, such that even under normal physiological conditions the medulla is close to hypoxia. Second, the tubular cells are highly metabolically active as their predominant function is sodium reabsorption through the active generation of an ATP-dependent diffusion gradient. Third, there is increasing evidence for the role of inflammatory mediators leading to direct tubular toxicity through complement- and neutrophil-mediated mechanisms.

Despite the label, only a small number of cells actually undergo cell death, and even then many of them appear to undergo apoptosis rather than true necrosis. The tubular cells, unlike podocytes within the glomeruli, have the fortunate ability to regenerate. Recovery occurs with the restoration of tubular cell number, a process which can take from several days to months, although, in the absence of repeated insults, recovery predominantly has occurred by 2–3 weeks.

In those who have undergone oliguric ATN, the return of tubular function may replace a small volume muddy brown urine with a large volume diuresis; this polyuric state may take several days to resolve as the tubular concentrating mechanism is re-established. During this period it is important to ensure that a further hypovolaemic renal insult does not occur.

Nephrotoxins

A variety of nephrotoxic medications (some of which are included in Table 10.4) are associated with intrinsic AKI, although the mechanisms by which they act are equally varied. NSAIDs, ACE inhibitors and angiotensin receptor blockers, as well as radiological contrast which causes a tubular necrosis, have already been discussed.

Aminoglycosides, such as gentamicin, are renally excreted and exert a direct cytotoxic effect on tubular epithelial cells, where they accumulate. Classically, a non-oliguric AKI occurs, and is related not only to the cumulative dose but also to the dosing regimen and duration of therapy. Aminoglycosides may be a contributory factor in AKI even after they have been discontinued because of drug accumulation. The risk of AKI can be reduced by the use of careful therapeutic drug monitoring. High-dose single daily dosing should be avoided in those at risk or with established AKI. Their use in patients with advanced renal impairment risks ototoxicity. Glycopeptide antibiotics have occasionally at high dosages been reported to cause nephrotoxicity; however, when used in conjunction with aminoglycosides they have a synergistic adverse effect and caution should be exercised in the use of these two classes concomitantly.

Other antibiotics, particularly penicillins and rifampicin, along with diuretics, proton pump inhibitors and NSAIDs are among a long list of potential drugs that can cause an acute inflammatory interstitial nephritis. This is a reversible hypersensitivity reaction to a drug or occasionally an infectious agent, which occurs 7 days to weeks after exposure. Clinically it may be associated with any combination of eosinophilia, arthralgia and a maculopapular rash and is treated by removal of the offending drug if it can be reliably identified; steroids are sometimes used. However, the incidence should not be overstated and should not preclude the use of these drugs in patients with AKI unless there are specific features of concern. Acute interstitial nephritis is unlikely to be the cause of AKI within 7 days of the start of the potentially offending drug.

Amphotericin, an antifungal agent, has a direct cytotoxic mechanism that leads not only to a decrease in GFR but also to profound hypokalaemia and hypomagnesaemia through renal losses into the urine. The effectiveness of this drug and also its toxicity relate to the cumulative dose. The introduction of liposomal formulations of amphotericin have reduced its organ toxicity, but at greater cost.

Sodium phosphate bowel preparations for colonoscopy have been implicated in a condition labelled phosphate nephropathy, which predominantly seems to occur in those patients with pre-existing CKD. The injury that occurs is not always reversible and this drug should be avoided in this group of patients.

Ironically, the immunosuppressant calcineurin inhibitors ciclosporin and tacrolimus, which are widely used for renal transplantation among other solid organ transplantations, can also be nephrotoxic. At high plasma concentrations they cause afferent arteriolar vasoconstriction and reduced GFR. Toxicity is commonly associated with hyperkalaemia. These effects can be reversed by dose reduction, but this should only be done after liaison with the responsible transplant team.

Rhabdomyolysis

Breakdown of muscle leads to a release of potentially nephrotoxic contents, and the inflamed muscle sequesters fluid causing volume depletion. Myoglobin is freely filtered at the glomerulus and free radical-mediated injury to the tubular cells is caused by the haem–porphyrin ring and obstruction through precipitation of myoglobin within the distal tubules. A wide number of aetiologies may lead to rhabdomyolysis, including muscle injury, drugs, toxins and infectious agents. However, in surgical disease it is often seen in the context of major trauma (particularly entrapment and burns), limb ischaemia and with an increasing frequency following surgery. Prolonged bariatric surgery in which there is high risk of pressure necrosis and also vascular surgery with prolonged use of tourniquet are implicated.

A clinical diagnosis can be made on the basis of characteristic dark 'tea-coloured' urine, as well as a plasma creatinine kinase level that is markedly elevated by several orders of magnitude. If an established oligoanuric AKI has developed then treatment can be difficult; however, if identified early, therapy is directed at maintaining a good diuresis. Large quantities of iso-osmolar saline are used, although the accumulation of oedema in muscle may result in greater quantities of infused fluid than is diuresed; volumes of >10 L/day are frequently required. Theoretically, myoglobin precipitates more easily at acidic urinary pH and intravenous sodium bicarbonate infusion keeping the urine pH >6.5 may reduce tubular precipitation. Hypocalcaemia results from calcium sequestration within damaged muscle; if, in addition, the patient is alkalinized by intravenous sodium bicarbonate, plasma ionized calcium levels may fall resulting in tetany, and the patient should be monitored carefully with this in mind. A forced diuresis is continued for several days and discontinued once creatinine kinase levels appear to be falling significantly.

Cholesterol embolization

Those patients with a high burden of atherosclerotic disease, usually the elderly, may develop an irreversible embolic AKI following vascular surgery, angiography, thrombolysis or occasionally spontaneously. Disrupted arterial plaque results in distal embolization that leads to vaso-occlusion and inflammation in the kidneys, the lower limbs, the brain and eyes and occasionally the gut. Characteristically, there is a rapid development of AKI following a procedure, associated with proteinuria and an eosinophilia and purpuric cutaneous infarcts over the lower limbs, particularly in the digits (described as 'trash feet'). There is no evidence that anticoagulation has either a primary or secondary preventative effect and may, if anything, increase the risk of the embolization of unstable plaque. Whereas contrast nephropathy and cholesterol embolization may supervene concomitantly and cause AKI following endovascular injury and intervention, the former normally resolves by 14 days, the latter persists and is a rare a cause of end-stage renal failure.

Multiple myeloma

Myeloma can cause renal impairment in a multitude of ways; however, it can occasionally present as an acute injury in the context of comorbid disease in the elderly. Volume depletion may lead to the precipitation of even modest levels of circulating paraprotein, which form insoluble casts within the renal tubules. The presence of anaemia and a detectable serum paraprotein are helpful in the diagnosis.

Oxalate nephropathy

Hyperoxaluria is associated with the formation of oxalate stones and also oxalate nephropathy, a progressive irreversible cause of renal impairment. Oxalate is found in a large number of plant-derived foodstuffs and, within the gut, it usually complexes with calcium, forming an insoluble precipitate that is egested in faeces. However, in the setting of malabsorption, free fatty acids saponify calcium and inhibit the formation of calcium oxalate complexes; free oxalate is then absorbed within the colon, resulting in hyperoxalaemia and hyperoxaluria. This is recognized to occur in conditions that are associated with fat or bile acid malabsorption such as inflammatory bowel disease, pancreatic insufficiency, bowel resection and blind loop syndrome as well as jejunoileal and Roux-en-Y bypass.

As well as forming macroscopic stones, less frequently a clinical picture similar to that seen in the rare congenital condition primary hyperoxaluria occurs, with renal tubular crystalline deposits of calcium oxalate leading to tubulitis, fibrosis and progressive renal impairment. The timing of oxalate nephropathy is less acute than other pathologies described here, but is an increasingly recognized complication of upper gastrointestinal surgery in which it may present as early as 4 months postoperatively. Following bariatric surgery, up to 25% of patients may have urinary oxalate levels sufficient to be associated with subsequent renal impairment. The clinical determinants of nephropathy remain unclear, but those undergoing bariatric surgery may be predisposed as a result of comorbidities of diabetes, hypertension and obesity in this group of patients.

Attempts to limit dietary oxalate may have some benefit; however, once renal impairment has developed there is often a rapid progression to end-stage renal failure. The role for early surgical reversal, when possible, of bypass procedures in delaying the progression of the disease is unclear.

Vasculitis

Systemic or renal limited vasculitis with an associated glomerulonephritis are uncommon causes of AKI but should be suspected in the setting of a clinical history of constitutional ill-health with weight loss and fevers, and organ-specific symptoms such as small joint arthralgia, haemoptysis, epistaxis, rash, serositis and neuropathy. The first presentation of a systemic vasculitis may occasionally be with abdominal pain and gastrointestinal haemorrhage. Regardless of the type of vasculitis, if a glomerulonephritis is present there will be evidence of renal inflammation on urinalysis with visible 'Coca-Cola' urine or microscopic haematuria, proteinuria and red cell casts on microscopic examination. Concern about AKI in the setting of vasculitis necessitates urgent discussion with a nephrologist about urgent specific immunosuppressive treatment.

Differentiating prerenal and intrinsic renal disease

A number of clinical and biochemical clues may be available to differentiate whether an unobstructed patient with AKI has prerenal or intrinsic renal disease, although, in practice, this is usually differentiated by the timing of the clinical course, with prerenal AKI responding rapidly to restoration of euvolaemia whereas an absent or prolonged response suggests intrinsic disease. Loss of urinary concentrating ability is an early sign of intrinsic disease with urine osmolality (U_{Osm}) below 350 mosmol/kg diagnostic, whereas a level greater than 500 mosmol/kg in the absence of diuretics demonstrates an appropriate concentrating response seen in prerenal AKI. The fractional excretion of sodium (EF_{Na}) is generally <1% in prerenal and >1% in intrinsic disease but correlates poorly in those patients on diuretics with CKD and may not hold true for contrast-induced AKI.

The early involvement of a nephrologist is important in intrinsic renal disease, particularly when there is a history and clinical evidence of renal inflammation, features of a multiorgan systemic disease and prolonged presumptive ATN. A renal biopsy will help determine the cause of an intrinsic AKI but in practice is reserved only for those in whom assessment suggests that there may be a response to disease-specific therapy as in acute interstitial nephritis or glomerulonephritis.

Care of the patient with acute kidney injury

Promoting recovery from acute kidney injury

The initial steps in management of AKI are to ensure restoration of renal perfusion with fluid resuscitation and vasopressors or inotropes when needed, discontinuing nephrotoxic medications, while excluding obstruction and considering other causes of renal impairment. Currently, there is no evidence for the use of any drugs to promote recovery of AKI.

Furosemide and other loop diuretics are widely used in advanced AKI, with the intention of converting oliguric to non-oliguric AKI. This may be perceived to be beneficial in selected patients as fluid balance becomes easier to manage but ultimately RRT may be indicated for other reasons. Furosemide exerts its action on active sodium pumps from the luminal side of the tubular cell, it is therefore important to increase the prescribed dose in patients with reduced GFR to ensure that there is a sufficient concentration of filtered drug in the tubular lumen. Furosemide reduces the metabolic demand of tubular cells and therefore may reduce cell injury when perfusion is poor. However, at least one meta-analysis has demonstrated no difference in outcomes relating to patient survival or need for dialysis. Furosemide is not without risks, most notably ototoxicity when used in large doses as is required in oliguric AKI.

Dopamine at low doses (up to 5 μg/kg/min) binds to the D_1 receptor, causing vasodilatation and increased blood flow to the renal, coronary and mesenteric arteries. Unfortunately, there is no evidence that it affects AKI outcome or mortality, but, similarly to the use of loop diuretics, it is associated with a diuresis. There is some evidence that atrial natriuretic peptide may be of some benefit but larger studies are required to demonstrate a consistent benefit.

General complications of acute kidney injury

The subsequent aspects of management of AKI are broadly to ensure appropriate clinical monitoring to allow prompt attention to complications of renal impairment, many of which are treated similarly if they arise from CKD. Monitoring should include daily, if not more frequently, assessment of circulating volume and weight and blood tests (renal function, haemoglobin, bicarbonate, calcium and phosphate).

Sepsis needs to be identified early and inflammatory responses examined for infective cause and treated aggressively. The anorexic and catabolic nature of renal impairment is often poorly matched by dietary intake and this needs to be addressed.

A variety of drugs will need modification in both delivery and dosing, especially those which are renally excreted. Of particular note in the surgical patient are analgesics and anticoagulants; gabapentin and all opioids may accumulate easily causing narcosis, equally heparins (including low-molecular-weight heparins) need to be dose reduced substantially. In those on dialysis and with advanced CKD there is no evidence for beneficial use of low-molecular-weight heparins for thromboprophylaxis and the risk of haemorrhage is increased by other uraemic factors such as gastrointestinal ulceration, angiodysplasia, anaemia and platelet dysfunction. Metformin and sulphonylureas can accumulate and should be stopped with AKI or advanced CKD. Rarely, penicillin-based and carbapenem antibiotics may also accumulate leading to neurotoxicity, with confusion, agitation and fitting.

Renal complications of acute kidney injury

Fluid overload is probably the most common problem associated with established AKI. In patients with positive fluid balance, large fluid intake and inadequate urine output, steps should be taken to limit fluid intake while avoiding intravascular volume depletion. Life-threatening pulmonary oedema is a particular risk in the oligoanuric patient in this setting; urgent treatment with topical or intravenous nitrates, non-invasive ventilation and occasionally loop diuretics may allow time for commencement of RRT. Pulmonary oedema is more likely in inflamed patients with increased capillary wall permeability and low intravascular oncotic pressure.

Rising serum potassium levels >5.5 mmol/L should prompt a removal of sources of oral or intravenous potassium supplementation. Overt hyperkalaemia >6.0 mmol/L may be associated with cardiac conduction defects and requires urgent treatment with intravenous calcium to stabilize the heart. Intravenous insulin with dextrose, sodium bicarbonate and nebulized salbutamol will all shift potassium to the intracellular space but have a duration of action of only hours and the clinician should be aware of the possibility of rebound hyperkalaemia. Hypertonic intravenous fluids may cause tissue necrosis with extravasation, and reliable venous vascular access should be used ideally via a central venous catheter. To avoid precipitation of calcium salts intravenous calcium and bicarbonate should never be administered via the same vascular access. Oral calcium-based resins that bind potassium are poorly tolerated and provide no benefit in acute hyperkalaemia and little benefit in the prevention of chronic hyperkalaemia by reducing dietary potassium absorption. Refractory hyperkalaemia >6.5 mmol/L in the setting of AKI is an indication for urgent dialysis.

The use of intravenous sodium bicarbonate in the treatment of persistent metabolic acidosis is contentious. It can be used in those patients with a modest acidosis that relates to AKI alone and is particularly useful in controlling hyperkalaemia. When AKI coexists with a lactate or ketoacidosis its use may lead to a paradoxical worsening of any acidosis. Progressive acidosis and uraemia are further indications for acute RRT.

Renal replacement therapy in acute kidney injury

A variety of dialytic modalities are available for supportive care in AKI. Local practice will determine whether they are delivered in a critical care setting or a nephrology unit; in those with multiorgan dysfunction, it is likely to be the former. Acute peritoneal dialysis is only very rarely used outside of resource-poor settings and will not be discussed further. The main divisions are between intermittent haemodialysis (iHD), continuous RRT (CRRT) and sustained low-efficiency dialysis (SLED). There is no strong evidence supporting the use of one therapy in preference to the others, apart from limitations of each technique in certain settings.

iHD differs little from standard haemodialysis offered in end-stage renal failure; it offers excellent solute clearance and fluid removal over a short timescale. This makes it ideal for stable patients with isolated AKI. However, the high ultrafiltration rates that are required can exacerbate haemodynamic instability and this precludes its use in a large number of patients with AKI. In reality, those patients requiring circulatory support with inotropes, vasopressors or intra-aortic balloon pumps tolerate iHD poorly.

CRRT encompasses a large number of techniques that utilize slow ultrafiltration rates. The major benefit is improved haemodynamic stability arising from slower blood flow rates. The disadvantage is poor solute clearance – these techniques need to be used for up to 24 hours a day to render them effective; also, interruptions due to clotted circuits or time off the machine can lead to a significant reduction in the delivered dose of renal replacement. One of the more widely used techniques is continuous venovenous haemofiltration (CVVHF), which uses, as does iHD, a semipermeable membrane across which a hydrostatic pressure is generated so that solute and fluid are removed by ultrafiltration. All the ultrafiltrate is discarded and a sterile replacement fluid is added to blood before returning to the patient in CVVHF. This is in contrast with iHD, which uses a countercurrent exchange mechanism to control the removal of fluids and solutes, such that the majority of fluid processed returns to the patient and no replacement fluid is required.

SLED lies as an intermediary technique between these two, utilizing the slow ultrafiltration rates of CRRT, coupled with the excellent solute clearance offered by iHD. The technique is gaining in popularity but is not widely available.

All forms of RRT require some form of vascular access; this is usually an uncuffed, temporary dual-lumen catheter placed in a femoral, subclavian or jugular vein. Both lumens need to be of large bore to allow the high flows of at least 250 mL/min required by these techniques. The major complications of catheters relate to problems at the time of insertion, and subsequent risk of infection and central venous thrombosis and stenosis, the latter is more common with the subclavian approach and has implications for patients who survive with end-stage renal disease requiring long-term

RRT when permanent vascular access may be compromised. Infection risk can be reduced by the use of tunnelled lines in those patients who may require ongoing RRT. The use of arteriovenous shunts has gone out of practice because of the high rate of arterial vascular complications.

Another practical consideration is the need for anticoagulation to prevent clotting within the extracorporeal circuit. The preference would be for an effective regional anticoagulant that was active only within the circuit; sodium citrate, heparin and prostacyclin (a short-acting potent platelet inhibitor) can all be used but invariably may have systemic effects if overused and result in clotting of the circuit if underused.

Future advances in acute kidney injury

Given that a rise in serum creatinine is a delayed marker of AKI and can have poor specificity, there are a number of novel biomarkers that are being investigated that not only may allow a more rapid identification of AKI but also be predictive of the severity of injury. More, specific markers of the nature of injury may allow the development of targeted therapies. Two such potential markers are neutrophil gelatinase-associated lipocalin, which appears early after AKI in both urine and serum, and cystatin C, which may have a role in identifying chronic as well as acute kidney disease. It is likely, however, that future developments may also rely on the profile of several markers to generate a 'profile' rather than relying on a single biomarker.

In the patient undergoing surgery and identified to be at high risk for AKI, remote ischaemic preconditioning may potentially have a protective role. Remote ischaemic preconditioning describes the phenomenon in which transient non-lethal ischaemia and reperfusion applied to one organ or tissue protects another organ or tissue from a subsequent episode of lethal ischaemia and reperfusion. This currently involves the use of a blood pressure cuff inflated to 200 mmHg for several minutes on an upper or lower limb either in the hours prior to surgery or after anaesthetic induction. There is some evidence from early small trials that it may reduce the incidence and severity of AKI in selected patients.

Chronic kidney disease

When the CKD staging is applied to data from the general population it is estimated that 10% of people have significantly advanced CKD. Therefore, this will commonly be the case in hospitalized surgical patients and, as has been outlined, increases a patient's risk of surgical mortality and morbidity including AKI. The following sections briefly outline the definition, associated complications and implications for surgical management.

Definition and epidemiology

Despite the nomenclature, CKD is not a single pathological process but represents a large number of pathologies that result in irreversible loss of kidney function. The term covers a spectrum within the population including early asymptomatic renal damage to patients with end-stage renal failure undergoing RRT.

A common staging system has been adopted worldwide in an attempt to aid both primary and secondary healthcare in the identification and prioritization of patients with early renal damage in order to target interventions that may delay the progression to end-stage renal failure and deal with the complications associated with CKD. For most patients, GFR is estimated (eGFR) by the use of formulae such as the Modification of Diet in Renal Disease (MDRD) formula, which incorporates serum creatinine and age, sex and ethnicity among others. The staging of CKD is based primarily on eGFR, with secondary determinants being the presence of significant proteinuria or structural kidney abnormalities with eGFR >60 mL/min/1.73m^2, when the formula is not so predictive of true GFR; Table 10.5 outlines the National Institute for Health and Clinical Excellence CKD classification.

Creatinine is an end-product of muscle catabolism that is released into the circulation at a constant rate and is freely filtered at the glomerulus as well as being secreted by the renal tubules. However, there is significant interpatient variability that relates to differences in muscle mass; the use of eGFR attempts to overcome these limitations. It is important to note that the 'normal range' for serum creatinine in laboratory results needs to be interpreted with caution: it is possible for a large, young black man with high muscle mass to have an 'abnormal' high creatinine with a normal GFR, while a small, elderly non-black woman may have a 'normal' creatinine yet have significant renal impairment.

CKD may be a progressive condition and it is generally accepted that patients will on average have a decline in eGFR by 1–2 mL/min/year. A number of factors are recognized to accelerate this decline, most notably diabetes, poorly controlled hypertension and proteinuria; the presence of these together can result in a deterioration of up to 15–20 mL/min/year. However, with good management, focusing on aggressive blood pressure control (<130/80 mmHg in a high-risk group), along with glycaemic control, smoking cessation and avoidance of nephrotoxic drugs, progression can be slowed.

Proteinuria is included as a secondary measure of CKD stage as it is the strongest single predictor of GFR decline in both the

Table 10.5 Kidney Disease Outcomes Quality Initiative (KDOQI) classification of chronic kidney disease

Stage*	GFR (mL/min/1.73m^2)	Description
1	≥90	Normal kidney function but urine or other abnormalities point to kidney disease
2	60–89	Mildly reduced kidney function; urine or other abnormalities point to kidney disease
3	30–59	Moderately reduced kidney function
4	15–29	Severely reduced kidney function
5	<15	Very severe or end-stage kidney failure

*The suffix (p) may be used to denote the presence of significant proteinuria.

diabetic and non-diabetic population. In the healthy glomerulus, protein is not filtered and remains within the capillary, although some low levels of protein are actively secreted in the tubules such as Tamm–Horsfall protein (uromodulin). Progressive renal damage is associated with glomerular scarring and an increase in protein leak at this site; proteinuria may itself perpetuate further renal damage. Large trials have demonstrated that the reduction of proteinuria by the use of drugs such as angiotensin receptor blockers and ACE inhibitors predict a better outcome. These drugs are therefore widely used in CKD not only for their beneficial effect on blood pressure but also for their renoprotective effect.

Proteinuria is defined as a urinary protein excretion of >500 mg/24 h; however, owing to their lack of reliability, the use of 24 hour urinary collections has been superseded by the use of either urine albumin–creatinine (ACR) or protein–creatinine (PCR) ratios. Using these methods, significant proteinuria is considered to be present when a urinary ACR >30 mg/mmol or PCR >50 mg/mmol or more is present. PCR >100 mg/mmol may indicate a glomerular pathology as opposed to or in addition to a tubular pathology. PCR >300 mg/mmol is considered to be in the 'nephrotic range', and may be associated with nephrotic syndrome, which may be due to membranous nephropathy and have an underlying surgical pathology such as a solid organ tumour. Nephrotic syndrome is also associated with a hypercoaguable state and risk of venous thromboembolism – highly relevant perioperatively.

Complications of chronic kidney disease

Sodium, potassium and water retention

At modest reductions in eGFR, the kidney is able to maintain sodium homeostasis but this is lost in almost all patients as renal function declines, resulting in sodium retention and extracellular volume expansion. This presents clinically with pulmonary and peripheral oedema, requiring the use of loop diuretics and sodium restriction to control. The relative insensitivity of the renal tubule to diuretics in advanced CKD means that high doses (up to 500 mg/day furosemide) may be required. Loop diuretics are also helpful in aiding potassium excretion, although the mainstay of therapy is dietary restriction. Avoidance of excessive volumes of parenteral fluids, particularly those containing significant quantities of potassium, is important in these patients perioperatively. Anaesthetic muscle relaxants such as suxamethonium and the catabolic status of the perioperative patient both predispose to hyperkalaemia. Postoperative assessment of potassium levels is strongly advocated in patients with advanced CKD.

Metabolic acidosis

Failure to adequately excrete hydrogen ions, while also losing the ability to prevent tubular bicarbonate loss, manifests as a metabolic acidosis, which further impairs potassium excretion. Symptomatic dyspnoea is unusual outside of advanced CKD, but oral sodium bicarbonate is used at earlier stages of renal impairment to correct the acidosis, which may aid in the management of hyperkalaemia, slow progression of CKD and reduce demineralization of bones.

Haematological

CKD is usually associated with reduced levels of erythropoietin (EPO), which is produced by interstitial fibroblasts within the renal cortex in response to hypoxia. EPO controls production of red blood cells within the bone marrow and the inadequate levels in CKD results in a normochromic normocytic anaemia. This is treated with the use of recombinant human EPO, while supplementary intravenous iron is usually also required to support erythropoiesis – inadequate gastrointestinal absorption hinders adequate oral supplementation.

Advanced renal dysfunction is associated with a coagulopathy that is not characterized by conventional laboratory techniques. In part this is related to platelet dysfunction, but there is also evidence of other coagulation factor anomalies. Apart from dialysis in those with advanced renal impairment, there are few specific treatments, although there is some evidence for the use of desmopressin but it is of limited efficacy. The surgeon should therefore anticipate that the patient with CKD will bleed, and certainly for planned surgery in patients who are dialysis dependent adequate preoperative heparin-free dialysis and optimization of haematocrit through blood transfusion may be desirable. Liaison with the nephrologist is valuable.

Cardiovascular

There is a close association between worsening eGFR and risk of cardiovascular events; even from the earliest stages of CKD, cardiac disease is the leading cause of mortality in end-stage renal failure. Furthermore, the presence of proteinuria is also independently associated with cardiovascular risk, and reduction of proteinuria in clinical trials reduces cardiovascular risk. Some of the increased cardiovascular risk with CKD is attributable to traditional vascular risk factors; however, there are a number of novel risk factors including endothelial dysfunction, anaemia and vascular calcification which are peculiarly important in renal patients. Renal patients have increased frequency of adverse cardiac events, left ventricular hypertrophy, stroke and peripheral vascular disease. Preoperative appreciation of these risks is important in deciding on intervention; however, quantitative assessment is extremely difficult. Patients should not be prejudiced from effective surgical treatment and involvement of the nephrologist in perioperative management is welcomed.

Immunity

After cardiac disease, the second most frequent cause of death among those patients with end-stage renal failure is infection. CKD is a chronic immunosuppressed state in which there is increased susceptibility not only to standard bacterial infections but also to reactivation of mycobacteria and a failure to clear chronic viral infections such as hepatitis B. There is *in vitro* evidence of impaired cellular and humoral responses in these patients. Poor nutrition is a common cofactor, and healing is impaired. Infection therefore needs to be avoided by timely removal of prosthetic materials; for example, urinary catheters and central venous catheters and infection should be treated aggressively and promptly with antibiotics.

Chronic kidney disease: metabolic bone disease

Metabolic bone disease is initiated by a reduction in phosphate excretion in the kidney and failure of 1a-hydroxylase in the kidney to produce circulating 1,25-dihydroxyvitamin D; these drive hyperparathyroidism. The consequences include osteodystrophy, a failure to buffer serum calcium levels and, importantly, soft-tissue calcification within the vasacalature. As a result patients with CKD have an increased fracture risk. Calcific arteriosclerosis leads to high rates of cardiac events, stroke and peripheral vascular disease. Stiffened vessels are less adaptive to haemodynamic insult and the consequences of shock are more profound. All these elements are important to take into account when assessing and managing surgical patients with renal disease.

Treatment of CKD metabolic bone disease in the earliest stage involves limiting phosphate intake both by dietary restriction and also by the use of phosphate binders – drugs that complex with phosphate in the gastrointestinal tract and prevent its absorption. Correction of metabolic acidosis and the use of activated vitamin D and calcimimetics have a role. At a later stage, described as tertiary hyperparathyroidism, when parathyroid adenomas have evolved from the hyperplastic glands of secondary hyperparathyroidism, control can be obtained only by surgical parathyroidectomy. Rarely, the latter is required as an emergency procedure for the treatment of calciphylaxis, a calcific vasculitis-associated tissue necrosis in hyperparathyroidism that carries a mortality of >50% if untreated. Patients with cutaneous ulcers associated with this condition may present to the plastic surgeon, and lack of recognition with local treatment in the absence of parathyroidectomy is futile.

Renal replacement therapy

Despite efforts to minimize the progression of CKD, many patients will progress to end-stage renal failure. Some patients' first presentation to a renal physician may be at end-stage, particularly if they have been relatively asymptomatic, not had contact with medical providers or CKD has not been identified. In those with end-stage renal failure, there are a variety of potential forms of treatment, the choice of which is a combination of patient preference and clinical indication. RRT may involve modalities of haemodialysis, peritoneal dialysis and transplantation; indeed, many patients will experience more than one of these modalities and sometimes all of these during their lifetime. However, supportive care with the choice not to initiate RRT or to discontinue established RRT may be recommended for those patients for whom it will not improve their overall quality of life and may even worsen it. In those who opt for RRT, there is no evidence for commencing dialysis on the basis of eGFR; practice varies but the indications include symptomatic uraemia, as well as hyperkalaemia and fluid management refractory to medical therapy. In those for whom transplantation is an option, pre-emptive transplantation prior to the requirement for a dialytic therapy is preferable.

Peritoneal dialysis

Peritoneal dialysis utilizes the patient's peritoneum as the dialytic membrane. Sterile fluid of volumes up to 2.5 litres is instilled by the patient or carer into the peritoneal space through a cuffed catheter sited on the anterior abdominal wall. This is then left for a variable period of time, during which solutes and fluid diffuse across the peritoneal membrane. The waste fluid is then drained out before a new bag is instilled. This is performed either as continuous ambulatory peritoneal dialysis, when three or four exchanges occur across the day, with an exchange left to dwell overnight, or as automated peritoneal dialysis when a machine cycles a number of exchanges overnight and a further exchange may be used during the day. The cuffed catheter can be inserted under local anaesthetic and may remain in position for many years. In general, dialysis is not started for several weeks until the breach in the peritoneum has healed and there is no evidence of a leak; the latter may present with perineal swelling, for example in the scrotum. Abdominal hernias require surgical attention, as the increased peritoneal pressure can rapidly exacerbate the size of the sac.

As a dialytic therapy it is used globally and has the advantage that it can be performed by the patient at home, at their place of work or while travelling, as long as they have an available supply of peritoneal dialysis fluid. The major complication is peritonitis, through the inoculation of bacteria or fungus into the peritoneum. The presentation is often of abdominal pain and is associated with cloudy waste drainage fluid, which demonstrates a leucocytosis on microscopy. Dialysis can usually continue while it is treated with intraperitoneal antibiotics. A more devastating but rare complication is encapsulating peritoneal sclerosis, a progressive fibrotic condition characterized by the coexistence of peritoneal sclerosis and features of bowel encapsulation with obstruction or altered function and is associated with a high mortality. The exact causation is unknown, but risk factors include length of time on peritoneal dialysis and episodes of bacterial peritonitis. Treatment is primarily conservative with the use of home parenteral nutrition if necessary. Peritonectomy is a surgical solution advocated in specialist centres for some patients with beneficial results.

Haemodialysis

In the stable patient, haemodialysis or haemodiafiltration utilizes a combination of diffusion and convection across an artificial extracorporeal membrane to remove waste fluid and solutes. This is performed traditionally on an intermittent three times a week basis in an outpatient dialysis unit but there is a resurgence in home haemodialysis, performed by the patient up to six times week. The length of each session varies between patients but is usually between 3 and 5 hours.

The success of haemodialysis is heavily dependent upon vascular access to allow blood flows of >350 mL/min required for effective dialysis. This can be achieved by the use of an arteriovenous fistula or an arteriovenous graft, which are then needled at each dialysis session or via an indwelling central venous catheter. Fistulas are the preferred form of dialysis access and are formed by the surgical anastomosis of an artery and vein, usually in the arm: radiocephalic and brachiocephalic and brachiobasilic. The native fistula requires a period of 4–6 weeks to mature, and failure of formation to function rates vary from 20% to 50%. A synthetic graft may be formed when there is unsatisfactory vasculature to form a fistula by interposing a synthetic tube graft that can be needled immediately; however, these are complicated by a high thrombosis and infection rate. Steal syndromes from distal tissues are occasionally seen with both fistulas and grafts. If dialysis is required as an emergency without other forms of access available, then a non-tunnelled central venous catheter may be placed; ideally, this should be converted to a tunnelled catheter at the first opportunity. The tunnelled catheter is typically placed in the internal jugular vein and is then tunnelled under the skin on the ipsilateral side; a profibrotic cuff over the external catheter tube secures it in place. While there is a risk of bacteraemia with a fistula, indwelling tunnelled catheters carry a greater risk of infection mitigated somewhat by the tunnel. The traumatic effect of the catheter on the endothelium of the central veins may also predispose in the long term to the development of a central venous stenosis.

Renal transplantation

For those patients in whom it is feasible, transplantation remains the pre-eminent from of RRT, with clear evidence of improved survival over dialysis in age-matched populations. Ideally, these patients are best served by a pre-emptive live donation as allograft survival from a living donor surpasses that seen with a well-matched deceased donor organ. A well-functioning transplant offers some protection against many of the complications of CKD and provides a more effective form of solute and fluid clearance than dialysis, as well as approximating far better the other physiological roles of the kidney. Moreover, there is a profound symptomatic benefit in those who undergo transplantation following dialysis, over and above the freedom from alternate daily or even more frequent medical interventions. However, the effect of CKD on cardiovascular risk and impaired immunity is not obviated by transplantation, and morbidity from these two groups of conditions remains high.

Patients with insulin-requiring diabetes mellitus may be considered for associated pancreas transplantation, predominantly type 1 diabetic patients with a smaller number of type 2 patients for whom islet failure has developed and insulin resistance is not overwhelming. The technique was pioneered through simultaneous pancreas–kidney transplantation, but there may be reasons for considering pancreas after kidney transplantation; for example, after live kidney donation. The relative benefits of whole organ pancreas and islet cell transplantation remains controversial.

GUIDE TO FURTHER READING

Augustine T, Brown PW, Davies SD, *et al.* Encapsulating peritoneal sclerosis: clinical significance and implications. *Nephron Clin Pract* 2009;**111**:149–154.

Borthwick E, Ferguson A. Perioperative acute kidney injury: risk factors, recognition, management, and outcomes. *BMJ* 2010;**341**:c3365.

Bosch X, Poch E, Grau JM. Rhabdomyolysis and acute kidney injury. *N Engl J Med* 2009;**361**:62–72.

Kellum JA, Unruh ML, Murugan R. Acute kidney injury. *Clin Evid (Online)* 2011;**ii**:2001.

McCullough PA. Contrast-induced acute kidney injury. *J Am Coll Cardiol* 2008;**51**:1419–1428.

Murugan R, Kellum JA. Acute kidney injury: what's the prognosis? *Nat Rev Nephrol* 2011;**7**:209–217.

Nasr SH, D'Agati VD, Said SM, *et al.* Oxalate nephropathy complicating roux-en-y gastric bypass: an underrecognized cause of irreversible renal failure. *Clin J Am Soc Nephrol* 2008;**3**:1676–1683.

National Confidential Enquiry into Patient Outcome and Death. *Acute Kidney Injury: Adding Insult to Injury.* London, UK: NCEPOD, 2009. Available from: http://www.ncepod.org.uk/2009aki.htm

National Kidney Foundation (United States). *Chronic Kidney Disease Guidelines.* Available from: http://www.kidney.org/professionals/kdoqi/guidelines_ckd/p1_exec.htm

Powell-Tuck J, Gosling P, Lobo DN, *et al. British Consensus Guidelines on Intravenous Fluid Therapy for Adult Surgical Patients.* Available from: http://www.renal.org/Libraries/Publications/British_Consensus_Guidelines_on_Intravenous_Fluid_Therapy_for_Adult_Surgical_Patients.sflb.ashx

Power A, Duncan N, Goodlad C. Management of the dialysis patient for the hospital physician. *Postgrad Med J* 2009;**85**:376–381.

Prowle JR, Bellomo R. Continuous renal replacement therapy: recent advances and future research. *Nat Rev Nephrol* 2010;**6**:521–529.

Thiruchelvam PT, Willicombe M, Hakim N, *et al.* Renal transplantation. *BMJ* 2011;**343**:d7300.

CHAPTER 11

Surgical infections and infestations

SIR ALFRED CUSCHIERI

Introduction

The biological factors involved in infection (infestation in the case of parasites) are:

- the virulence of the invading microbe
- the size of the infecting inoculum
- the state of the host's defence system.

At its most basic level, infection represents an adverse shift of balance in favour of microbial invasion that overwhelms the host defence system. In immunologically competent individuals the invasion is determined by the product of the virulence and the size of the inoculum of the pathogenic organisms. In immunocompromised patients low-virulence organisms and even commensals that are not normally considered pathogenic can obtain a foothold and cause serious infections.

Classification and consequences of infections

From a practical standpoint infections in surgical patients are best categorized as:

- community-acquired infections
- nosocomial (hospital)-acquired infections that cover (1) common postoperative infections and (2) infections in critically ill patients, usually in intensive care units (ICUs).

This distinction is important for various reasons, e.g. varying spectrum of infectious disease within the two settings, frequent presence of immune deficiency in hospital patients and significantly higher prevalence of infections by antibiotic-resistant organisms in hospital-acquired infections.

Nosocomial (hospital-acquired) infections afflict 3–5% of patients and are the most important contributors to prolonged hospital stay, increased costs and death after surgical treatment. On average, a postoperative wound infection increases the costs of an operation by 300–400%. Infections that arise following discharge from hospital are not reported consistently and probably account for half of the infections after surgical treatment. These postdischarge infections have assumed greater importance to the overall assessment of surgical morbidity and costs with the significant increase in ambulatory and short-stay surgical care. A minimum of 45 days of follow-up is necessary to obtain a valid assessment of wound infection rates.

Community-acquired infections cover a wide range, with some being primarily medical in the first instance, e.g. respiratory infections, meningitis, acquired immunodeficiency syndrome (AIDS) and viral hepatitis, and others surgical, e.g. abscesses, appendicitis and peritonitis. Thus, surgical community-acquired infections require emergency surgical treatment.

Consequences of nosocomial infections

The importance of nosocomial infections is highlighted by the statistic that they are the sixth leading cause of death in the USA, and hospital surveys indicate a rising incidence in most Western countries, with many being caused by antibiotic-resistant bacteria. In addition, nosocomial infections incur added costs and increase the discomfort and disability experienced by patients following elective operations. For patients who sustain serious injury and those who require critical care, infection is frequently lethal. The onset of bacterial or fungal infection contributes, as an independent causative factor, to the development of multiple organ systems failure.

Postoperative surgical infections

Infections following surgical operations fall into three categories:

1. systemic: including postoperative chest and urinary infections
2. surgical site: infection at the operative site (usually abscess formation)
3. wound infection.

It is important to stress that around 70% of postoperative infections present in the community after discharge from hospital.

Postoperative pyrexia

Pyrexia is a common feature of postoperative infections, although it may be absent in immunologically compromised patients. Its important features are:

- time of onset
- degree of pyrexia and type (persistent, intermittent)
- accompaniments, particularly rigors (shivering) and haemodynamic change.

Most early postoperative fevers are due to non-infectious causes, particularly pulmonary collapse. Early wound infections (occurring within 18–24 hours) tend to be of a serious nature (see below). Rigors indicate bacteraemia or viraemia and always necessitate blood culture in addition to physical examination. A flushed appearance with a hyperdynamic circulation is also indicative of bacteraemia and usually signifies the early stages of septic shock. Intermittent pyrexia is indicative of an abscess.

If a patient develops a high temperature in the postoperative period, the following are necessary in all patients: physical examination of the lungs, wound, calves and urine. In addition, most would advise a chest radiograph. Leucocytosis is a feature of the early postoperative period when cultures are usually negative. Thus, these laboratory tests (white blood cell count, total and differential) and culture screen (sputum, urine, blood and wound) are undertaken on a selective basis and in response to specific clinical findings:

- fever that persists beyond the first 24 hours
- fever that recurs after a period without fever (intermittent)
- fever that arises after the first 24 hours
- fever accompanied by rigors, haemodynamic change or chest/abdominal signs.

Procurement of material for culture

This is an important practical aspect of management as it ensures valid culture results and aids recovery of organisms that are difficult to culture. The first principle is to retrieve an uncontaminated specimen in a sterile container that does not contain air if anaerobes are thought to be involved in the infection. Thus, material from intra-abdominal abscesses that often contains anaerobic organisms should be aspirated into a sterile syringe and all the air expelled. The specimen should then be transferred to an anaerobic transfer container or left in the syringe for direct inoculation into culture medium in the bacteriology department.

Gram staining of the infected material provides an important immediate evaluation while the results of culture are awaited. The request form accompanying the specimen should give sufficient details and, in particular, indicate the pathogens suspected whenever possible. Blood cultures should be obtained in duplicate, 20–30 mL per set obtained from different sites in adults. The skin overlying the venepuncture site should be scrubbed with antiseptic (povidone iodine) for 2 minutes beforehand and sterile gloves worn to prevent contamination. The wound surface should be cleaned with sterile water (not bacteriostatic saline) before obtaining the specimen, which is best obtained by scraping. Sometimes a small biopsy is needed for culture. Swabs of wounds or fluids are undependable, as they may not represent the fluid content of the wound accurately. Bullous lesions are aspirated after gentle surface cleaning. Body fluids (ascites, pleural effusions) are aspirated into a syringe after disinfection of the wound surface with povidone iodine. In some situations, quantitative cultures are needed, e.g. clean catch specimens of urine and burn sepsis.

Common postoperative infections

The common postoperative infections include:

- wound infections
- pneumonia
- urinary tract infections
- central venous line infections
- intra-abdominal abscess
- infection of implanted material.

Urinary tract infections

These are most commonly associated with catheterization of the bladder (indwelling and 'in and out'). Symptoms include dysuria and frequency and, sometimes, the onset of incontinence. Loin pain and tenderness are only found in patients with severe upper urinary tract infections. The specific diagnosis of urinary infection is made with the recovery of more than 10^5 organisms per millilitre of urine. The most common organisms cultured in nosocomial urinary infections are *Escherichia coli, Pseudomonas aeruginosa* and coagulase-negative *Staphylococcus* spp.

Central venous line infection

This is encountered in 3–5% of patients who have monitoring lines in place and those receiving parenteral nutrition. Fever in such patients is an indication for inspection of the puncture site for signs of inflammation and for changing the line. For temporary lines, removal and replacement to another site are indicated. Replacement over a guide wire may permit preservation of a valuable access site. Permanent lines may occasionally be salvaged, temporarily, by a course of systemic antibiotics. However, if signs of infection have not resolved completely within 48 hours, the catheter should be removed. The tip of all removed catheters should be cultured. *The presence of more than 15 colonies is indicative of line infection*. Episodes of line sepsis can be prevented by meticulous attention to cleanliness of nutrition support lines.

Intra-abdominal abscess

Intra-abdominal abscess can arise in:

- Patients undergoing a major operation involving the alimentary tract. The abscess results either from contamination at the time of surgery or from suture line leakage.
- Critically ill patients from failure of intraperitoneal host defences and the gut barrier function with translocation of intraluminal bacteria into the peritoneal cavity (sometimes referred to as tertiary peritonitis).
- Patients with acute abdominal conditions requiring emergency operations, e.g. trauma, perforated viscus or severe gangrenous perforated appendicitis. These patients have a 6–10% risk of developing intra-abdominal abscesses.

Signs and symptoms of postoperative abdominal abscess usually arise between the fifth and 10th postoperative day and include intermittent fever, localized tenderness and absent bowel sounds. In some instances the abscess is palpable abdominally or rectally (pelvic collection). Persistent drainage from an abdominal wound infection that has been opened indicates that this is being fed from an intra-abdominal site. Patients with intra-abdominal abscess may develop signs of sepsis such as hypotension, hyperdynamic circulation, respiratory distress and other features of the multiple organ failure/systemic inflammatory response syndrome (SIRS).

Plain radiology may show an elevated immobile hemidiaphragm in patients with subphrenic abscess and visible gas may be seen, especially in patients who have sustained anastomotic dehiscence or have developed a pancreatic abscess, but diagnosis is nowadays based on ultrasound scanning and especially CT. Both are usually performed, one after the other. CT provides more detailed information on the precise location and anatomy of the abscess cavity. CT-guided drainage is now used as the first line of treatment in these patients, with surgery being reserved for large multiloculated abscesses containing a large amount of slough. Systemic antibiotics are indicated to forestall the systemic effects of bacteraemia that may occur before, during or after the abscess drainage.

The key issue relates to the role of laparotomy in septic patients with SIRS and multiorgan failure. In the past, this was recommended in all these patients on the premise that the detection and evacuation of abscesses was followed by improvement or reversal in 50% of patients. This view is now not generally accepted, as the majority of patients with SIRS do not have an intra-abdominal focus of infection. The current consensus is that laparotomy is used selectively in patients who are in the early stages of the disease, especially those in whom the organ failure was precipitated by intra-abdominal infection in the first instance, i.e. detection of residual collections. Laparotomy is not indicated in patients in the late hypodynamic decompensated stage of the disease unless there are specific signs of intra-abdominal infection.

Infection of implanted prosthetic material

Infection is an ever-present risk when prosthetic material or implants are used; whether this is simple mesh for hernia repair or a more complex implant, e.g. joint prosthesis, vascular grafts, cardiac valves or pacemaker devices. The clinical evidence of infection of implanted materials ranges from subtle to catastrophic. Thus, implanted intravenous lines may exhibit only fever and local erythema at the puncture sites, whereas infected aortic grafts may present with life-threatening haemorrhage due to anastomotic disruption or acute aortoenteric fistula.

In nearly all instances, removal of the implant is necessary to control infection. Removal and replacement of implanted devices such as pacemakers to another site with systemic antibiotic cover is needed. For infected vascular grafts removal with a new prosthesis tunnelled through uncontaminated tissue (extra-anatomical bypass) is usually performed, e.g. axillobifemoral grafts for infected abdominal aortic grafts. In some patients, the threat to life or anticipated disability may prohibit removal of the implant, e.g. infected thoracic aortic graft. Irrigation of the infected area with antibiotic solutions in addition to systemic antibiotics may buy time but does not provide permanent control of the infection.

A growing percentage of nosocomial infections are associated with medical devices. One solution to this problem is to coat the device with anti-infective coating. Antibiotic coating has been used especially for vascular grafts. In this instance the polymer is impregnated with gelatin, which then binds antibiotics such as rifampicin. This approach has a number of limitations. In the first instance, each antibiotic is effective only against certain bacteria. Second, there is the problem of bacterial resistance and, third, antibiotic coating tends to be effective for only short periods (days to weeks). A more recent technology involves *oligodynamic iontophoresis*. This is an electrochemical process whereby minute amounts of silver ions are released from the device into the surrounding tissue fluid, thereby preventing bacterial colonization and infection of the implant. Oligodynamic iontophoresis-enhanced materials are polymers impregnated with silver, platinum and carbon particles. These components, following implantation of the device, set up an electrochemical reaction once in contact with the crystalloid tissue fluid, with the silver and platinum particles acting as the electrodes of a battery. The result of this reaction is the release of bactericidal silver ions that lasts for several months. Currently, oligodynamic iontophoresis-enhanced central line catheters are being evaluated in randomized trials.

Principles of antibiotic therapy

Antibiotics are rarely used as the sole agents to eradicate surgical infections; in most instances they constitute adjuvant treatment to surgical and radiological interventional procedures, e.g. excision of the infecting focus drainage of abscesses, debridement or lavage of infected serous cavities. They are certainly no substitute for effective surgical management of these disorders. The use of antibiotics as prophylactic agents to cover certain operations is well established and of proven value. *The adverse effects of antibiotics, particularly the emergence of resistant strains of organisms, limit their overall usefulness, especially in critically ill patients.*

Antimicrobial resistance

Over the past 20 years antimicrobial resistance (AMR) has become a major worldwide health problem with adverse effects on the treatment of a wide spectrum of infectious disease from tuberculosis to malaria and both community- and hospital-acquired infections. In certain countries such as the UK, meticillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Clostridium difficile* infections have reached epidemic proportions with regular outbreaks resulting in ward closures. In the European Union (EU) countries, 25 000 patients die each year from infections caused by drug-resistant bacteria with related costs exceeding €1.5 billion in healthcare expenses and productivity losses. Both the World Health Organization (WHO) and the European Commission have issued comprehensive action plans needed to address the problem of increasing AMR.

The EU Action Plan covers seven areas where measures are most necessary:

1. making sure antimicrobials are used appropriately in both humans and animals
2. preventing microbial infections and their spread
3. developing new effective antimicrobials or alternatives for treatment
4. co-operating with international partners to contain the risks of AMR
5. improving monitoring and surveillance in human and animal medicine
6. research and innovation
7. communication, education and training.

The mechanisms underlying AMR are complex and incompletely understood but in the final analysis the problem relates to the misuse or inappropriate use of antibiotics, which are grossly overprescribed in both primary and hospital healthcare. AMR is caused by mutation in plasmid (transferable) or chromosomal genes. Mutations remain at a low level in the population of bacteria under natural conditions but increase with stress induced by antibiotic exposure. Resistance genes can be transmitted between bacteria by transformation or conjugation. Resistance can occur to a single antibiotic or to several.

Enzymes

β-Lactam resistance in bacteria is the result of the expression of several classes of β-lactamase enzymes. This accounts for resistance to many β-lactams, including the third-generation cephalosporins (cefotaxime and ceftazidime) by several species of Enterobacteriaceae and *P. aeruginosa*. New plasmid-mediated β-lactamases, extended-spectrum β-lactamases, confer resistance to most β-lactams antibiotics except cephamycins and carbapenems. The genetic information encoding the β-lactamase can be transferred between many Gram-negative and Gram-positive bacteria.

Altered target site

Fluoroquinolone antibiotics exert their antibacterial effect by inhibition of DNA gyrase (bacterial topoisomerase II) and topoisomerase. Mutation in *gyrA* (encodes A subunit of gyrase) occurs predominantly in Gram-negative bacteria, whereas mutation in *grlA* (topoisomerase IV subunit A) occurs predominantly in Gram-positive bacteria. Double or multiple mutations within *gyrA* confers high-level resistance to fluoroquinolone antibiotics.

Diminished uptake and efflux of antibiotic

The concentration of antibiotic in bacteria can be reduced by decreasing the penetration (uptake) or by increasing efflux of antibiotic from the bacterial cell. It is thought that the major route through the bacterial cell membrane for hydrophilic antibiotics (fluoroquinolones, chloramphenicol, tetracycline and some β-lactams) is through water-filled protein pores called porins mediated through a protein OmpF, complete loss of which gives rise to cross-resistance to a number of antibiotics.

Increased efflux is another recognized mechanism of antibiotic resistance. This efflux is mediated by several export proteins, including the Bmr from *Bacillus subtilis* and the Tet efflux proteins and *NorA* gene of *S. aureus*. Bmr mediates the efflux of chloramphenicol and fluoroquinolones.

Transformation

This mechanism involves the uptake of naked DNA by bacteria from the environment. The most studied species include *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Neisseria* spp. Most naturally transformable bacteria usually take up DNA from a closely related bacterium, usually from DNA released by bacterial lysis.

Conjugation

This requires contact between two bacterial cells, thus enabling bacterial gene transfer. The genes are transferred either as plasmids or a conjugative transposon (designated Tn plus a number). Conjugation is thought to be the main mode of transfer of antibiotic resistance genes between bacteria. Large plasmids bearing several individual genes that each confer resistance to a different antibiotic are well documented, especially during prolonged antibiotic therapy.

Antibiotic policy

Certain principles governing antibiotic therapy in hospital practice are agreed and in general usage:

- Each hospital has its own drug formulary that includes an antibiotic policy (first-line antibiotics to be used for specific conditions) based on cost, efficacy, the pharmacokinetic properties and the hospital's known resistant species. This policy covers both the treatment of established infections and the use of specific antibiotics for prophylaxis of infection in patients undergoing surgery. By agreement, certain antibiotics are kept in reserve for serious infections.
- For established infections, the sensitivity of the organisms cultured to antibiotics is performed routinely and the first-line antibiotic regimen used may need to be changed accordingly.
- For certain antibiotics, therapeutic drug monitoring is necessary to (1) establish adequate serum concentrations and (2) identify potentially lethal concentrations. This applies particularly to aminoglycosides

(gentamicin, netilmicin, tobramycin and vancomycin) and flucytosine. The desirable levels of aminoglycosides vary according to the nature and severity of the infection. Dose adjustment is essential in patients with renal impairment, when advice should be sought from the clinical pharmacist. The volume of distribution of aminoglycosides is increased in critically ill patients and, for this reason, suboptimal dosing is common.

- In some infections synergistic combinations of antibiotics are indicated. Antimicrobial synergy occurs, for example, when an aminoglycoside is combined with penicillin for treatment of certain staphylococcal or enterococcal infections, and with ticarcillin for enhanced activity against *Pseudomonas* spp. For surgical patients, the traditional treatment of potentially life-threatening infections, e.g. pneumonia, suppurative cholangitis, peritonitis and burn sepsis, has been with combinations of aminoglycosides and other drugs such as cephalosporins, clindamycin and metronidazole.
- Infected collections negate antibiotic activity owing to changes in tissue pH, oxygen tension, levels of magnesium and calcium, and the production by various organisms of substances that inactivate antibiotics, such as β-lactamase, which inactivates penicillin. Thus, drainage and debridement will improve antibiotic effectiveness as well as reduce the bacterial inoculum.
- In serious infections in critically ill patients, discussion with and advice from the hospital clinical bacteriologist is essential.
- Special nursing measures and isolation of patients with MRSA infections are necessary.

Additional measures for enhancing antibiotic action include:

- altering body fluid pH, e.g. urine
- delaying excretion of the drug, e.g. use of probenecid with penicillin
- changing the route of administration, e.g. intravenous from oral
- increasing the dose of antibiotic (effective in cephalosporins); this can be achieved by increasing the absolute dose or the total dose (increase duration of therapy), or by reducing the dosing interval
- using substances that block bacteria-inactivating enzymes, e.g. clavulanic acid.

Types of antibiotic

Penicillins

The use of penicillins has been eclipsed by other antibiotics because of the emergence of resistant organisms that produce penicillinase, as well as the emergence of MRSA. Penicillins as a group are effective against Gram-positive organisms and *Neisseria gonorrhoea*. The aminopenicillin group (ampicillin, amoxicillin) also has limited Gram-negative activity. Penicillinase-resistant penicillins are useful against resistant staphylococci.

More recent penicillin classes include the carboxypenicillins (carbenicillin, ticarcillin) and ureidopenicillins (mezlocillin, azlocillin and piperacillin). The main advantage of carboxypenicillins is their effectiveness in *Pseudomonas* spp. and *Proteus* spp. infections. Ureidopenicillins are effective against *Pseudomonas* and *Klebsiella* spp. Resistance of β-lactamase-producing *Staphylococcus* spp. and *H. influenzae* is common. Reports of decreasing activity of ureidopenicillins with increasing inoculum size are important.

As a group, penicillins may produce allergic reactions that range from rashes (relatively common) to anaphylaxis (rare). Ampicillin is associated with diarrhoea. Haemolytic anaemia, drug fever, granulocytopenia and hepatitis are unusual but important side effects.

Cephalosporins

These were first introduced in the 1950s and are usually classified into three generations:

- *first-generation cephalosporins*: cefadroxil, cephalexin, cephradine, cefazolin, cephalothin, cephapirin
- *second-generation cephalosporins*: cefaclor, cefamandole, cefonicid, ceforanide, cefotetan, cefoxitin, cefuroxime
- *third-generation cephalosporins:* cefoperazone, cefotaxime, ceftazidime, ceftizoxime, ceftriaxone.

First-generation drugs are available in both oral and parenteral forms. Cefazolin has been used widely as a prophylactic drug in high-risk elective operations. Only one second-generation cephalosporin is available for oral use (cefaclor); this is active against *H. influenzae.*

The major advantage of second-generation cephalosporins is improved activity against Gram-negative organisms such as *E. coli* and *Proteus* spp. Activity against *S. pneumoniae* and *Streptococcus pyogenes* is equal to the first-generation drugs. Cefoxitin is a useful drug for surgical patients because of its activity against *Bacteroides fragilis* and persistent high tissue levels lasting for 3–4 hours. The level of activity against anaerobes is shared by cefotetan. Cefoxitin is useful as a primary drug directed against suspected mixed infections within the peritoneal cavity and is commonly used in trauma patients.

Third-generation cephalosporins have improved activity against Gram-negative organisms. Thus, some of the drugs in this category have been suggested as primary single-drug therapy for difficult infections such as nosocomial pneumonia and peritonitis. The major advantage of these drugs over combinations of aminoglycosides with earlier generations of cephalosporins for nosocomial infections is the lack of toxicity and lack of need to monitor drug levels. However, clinical studies have not documented a superior outcome in pneumonia or peritonitis when third-generation cephalosporins are compared with combination therapy with aminoglycosides. Toxic side effects of third-generation cephalosporins are unusual. Moxalactam is associated with prolongation of the prothrombin time.

Aminoglycosides

This group has a wide spectrum of activity against a large number of organisms and has consistently been shown to be useful in many difficult clinical infections, particularly pneumonia and peritonitis. Toxicity is the major drawback, involving renal damage and ototoxicity, the latter producing deafness and vestibular dysfunction. Gentamicin, tobramycin and amikacin are most often used. The drugs are given parenterally because of poor intestinal absorption.

Serum drug-level monitoring is essential to determine effective tissue levels of the drug. Timing of the blood-level assay is important. One level should be drawn within 30 minutes of beginning an infusion of drug, with the second (peak) level being drawn 30 minutes after the end of the infusion. Peak levels above 5 μg/mL are associated with improved survival from serious infections. With pneumonia, higher levels in the range of 10 μg/mL may be required.

Clindamycin

This antibiotic is useful in infections caused by *B. fragilis.* The drug suppresses protein synthesis by bacterial ribosomes and binds to the same subcellular site as chloramphenicol. Thus, simultaneous use of these two drugs is highly inadvisable. Clindamycin-associated diarrhoea, rarely progressing to pseudomembranous colitis (PMC), is an adverse side effect.

Vancomycin

This has assumed major importance in the treatment of surgical infections due to resistant micro-organisms, particularly MRSA and *Staphylococcus epidermidis* and *Corynebacterium diphtheriae.* In MRSA infections, the addition of aminoglycosides or rifampicin may be necessary (sometimes all three). Vancomycin should be administered over a 1 hour period to avoid hypotension and a red rash over the upper body (red man syndrome). Vancomycin is also used in the treatment of PMC caused by *C. difficile.* Resistance to vancomycin is now well documented among nosocomial infections.

Carbapenems

This group, which is related to the penicillins, is active against a variety of Gram-negative organisms and anaerobes. The most commonly used antibiotic of this group is imipenem–cilastatin. Cilastatin is a toxicity-free additive that inhibits the dehydropeptidase enzyme that catalyses the reaction leading to rapid excretion of imipenem by the kidney. Another member of the group is meropenem. These antibiotics are valuable in the treatment of serious abdominal and pulmonary infections caused by Gram-negative bacteria.

Monobactams

The only member of this group that is used clinically is aztreonam, which exhibits marked activity against aerobic Gram-negative organisms. Studies have confirmed outcomes equivalent to combination therapy with aminoglycosides in infections due to *E. coli, Klebsiella pneumoniae, Streptococcus marcescens, P. aeruginosa, Enterobacter* spp., *Proteus* spp. and *Providencia* spp. Combined with clindamycin, this drug has been effective in the treatment of aspiration pneumonia.

β-Lactamase inhibitors

The binding of the β-lactamase inhibitor clavulanic acid to certain antibiotics, e.g. ticarcillin, enhances their activity. Co-amoxiclav (amoxicillin + clavulanic acid) is effective against aspiration and postoperative pneumonia and peritonitis.

Quinolones

This group includes ciprofloxacin and norfloxacin. These drugs have a wide spectrum of activity against Enterobacteriaceae and may be administered orally and parenterally. Occasional nausea, vomiting and diarrhoea represent the most common side effects.

Metronidazole

Originally introduced for the treatment of protozoal infections (*Trichomonas vaginalis, Giardia lamblia, Entamoeba histolytica*), this is the most effective agent against anaerobic Gram-negative infections, *Bacteroides* and *B. fragilis,* and is often used in combination with cephalosporins and aminoglycosides for mixed infections. Metronidazole is also effective in PMC caused by *C. difficile.*

Infections in critically ill intensive care unit patients

Patients in ICUs are two to five times more likely to develop nosocomial infections than the general hospital population (where 5% prevalence is regarded as the norm). The acknowledged risk factors include:

- age >70 years
- shock
- steroids
- chemotherapy
- ICU stay >3 days
- previous antibiotics
- mechanical ventilation
- invasive monitoring
- indwelling urinary catheter >10 days
- acute renal failure
- surgical versus medical patient.

Invasive monitoring poses a real dilemma and the value of the widespread use of the pulmonary flotation catheters has been questioned recently. Selective use only in patients who require this invasive cardiac monitoring is a sensible option. The consequences of infection in ICU patients are always serious, with an established association with multiple organ systems failure and a fatal outcome.

Detection of infection in intensive care unit patients

Evaluation of the critically ill patient for infection is often difficult. The two problems are (1) differentiation between colonization and invasive infection and (2) localization of the site of infection. In the search for infection, a systematic head-to-toe approach is needed. The prior condition of the patient is also taken into consideration, i.e. pre-existing medical diseases and the surgical procedures that have been performed. Frequently, the occurrence of infection in ICU patients is closely associated with complications of surgical interventions. In this context, the nature of the operation is important, e.g. anastomotic dehiscence and intra-abdominal abscesses after gastrointestinal surgery, mediastinitis after cardiac surgery. Other hospital-acquired infections must also be excluded:

- *head and neck*: sinusitis, meningitis, ventriculitis, etc.
- *chest*: aspiration, pneumonia, lung abscess, empyema
- *abdomen*: *C. difficile* PMC, urinary tract infection
- *catheter-related infections*: arterial catheters, central venous catheters, Swan–Ganz catheters.

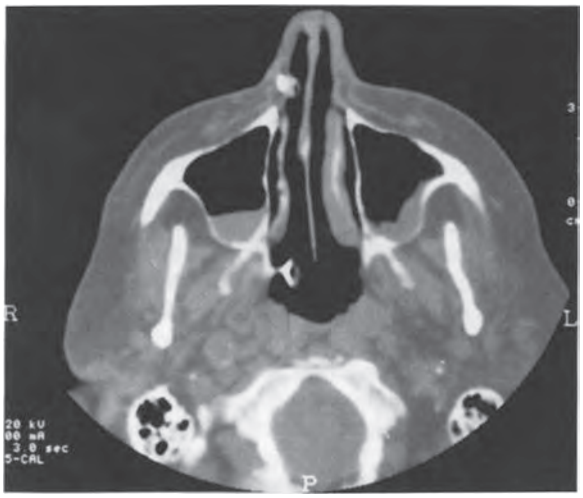


Figure 11.1 CT scan of the head in a patient who was septic from sinusitis. Note the air–fluid levels in the maxillary sinuses.

Sinusitis is most commonly associated with nasotracheal intubation but may also complicate indwelling nasogastric tubes. The overall reported incidence in ICU patients is 2%. Facial trauma is a predisposing factor. The maxillary sinuses are involved in 50–75% of cases. The mechanism by which the sinusitis occurs is usually obstruction of the sinus ostia by the tube; hence, the sinus involved is almost always on the same side as the nasal tube. The diagnosis, once suspected, is established by a CT scan (Figure 11.1). The bacteria involved are usually Gram-negative bacilli, *S. aureus* or anaerobes. Treatment includes removal of the nasal tube, decongestant sprays, antibiotics and drainage of the sinus.

Meningitis and ventriculitis are most commonly related to intracranial surgery, head trauma or the placement of intracranial pressure monitors. The most common organisms causing these infections are Gram-negative bacilli, *Staphylococcus* spp. and *Streptococcus* spp. The use of prophylactic antibiotics in patients with intracranial monitoring devices remains controversial.

Chest infections in intensive care unit patients

These are very common and include pneumonia, lung abscess, empyema and mediastinitis (after cardiac surgery). Pneumonia is by far the most common. Lung abscess usually arises against a background of pneumonia, and empyema is most commonly encountered in trauma patients with chest tube drainage and oesophageal perforation or leaks. The majority of ICU-acquired pneumonias reflect the ICU's bacterial flora. Within a short time of arriving in the ICU, the patient's own bacterial flora is replaced by the ICU flora, which often contains antibiotic-resistant strains. The risk factors for pneumonia are:

- age
- aspiration
- head injury
- smoking
- intubation and mechanical ventilation
- lung injury: acute respiratory distress syndrome (ARDS) and pulmonary contusion
- prior use of antibiotics
- prolonged preoperative hospital stay
- pulmonary oedema
- use of H_2 blockers and proton-pump inhibitors
- upper abdominal and chest incision.

Patients on mechanical ventilation are particularly at risk as normal protective mechanisms such as the mucociliary clearance mechanism and coughing are breached. The mechanism by which pneumonia occurs in ICU patients is thought to be by aspiration of upper airway and oropharyngeal secretions in most instances, as evidenced by the bacteriology. Thus, approximately 70–75% are caused by aerobic Gram-negative bacilli, 15–20% by *S. aureus* and 5–15% by *Candida*. Only 2–15% of normal humans have Gram-negative bacilli colonizing their upper airways, but, in critically ill patients, the colonization of the upper airway occurs within the first few days of ICU stay in 55–75% of patients.

The diagnosis of pneumonia in critically ill patients may be difficult. Chest radiograph may not be helpful as many of these patients already have pulmonary infiltrates from other causes, e.g. ARDS or pulmonary contusion (trauma patients). Thus, diagnosis is based on (1) quantity and quality of sputum, (2) Gram staining and (3) sputum culture. The presence of copious purulent sputum containing leucocytes and intracellular bacteria on Gram staining is enough to make a diagnosis pending the results of culture. Initial therapy is based on sputum Gram staining as follows:

- *Gram-negative bacilli*: third-generation cephalosporin, extended spectrum penicillin, monobactam, aminoglycoside
- *Gram-negative bacilli with suspicion of* Pseudomonas: imipenem or cilastatin, penicillin + aminoglycoside, third-generation cephalosporin + aminoglycoside
- *Gram-positive cocci in clumps:* nafcillin, vancomycin for MRSA
- *Gram-positive cocci in chains*: penicillin G
- *yeast:* amphotericin B, fluconazole.

The results of sputum culture may not be representative of the infection because of contamination by oral flora. To overcome this, samples may be obtained by transtracheal aspiration through the cricothyroid membrane via an Angiocath (in non-intubated patients) or via a flexible bronchoscope with a protected brush technique for sputum sampling.

Catheter-related infections

Patients in the ICU are more likely to develop catheter-related sepsis than those in hospital wards. However, it is unclear whether this is due to the patient's underlying illness or to the fact that these patients tend to have catheters that are often inserted under emergency conditions, are *in situ* for long periods and are frequently manipulated. Catheter-related bacteraemia is encountered when the number of colony-forming units cultured from the tip is 15 or more. The changing of catheters to prevent this bacteraemia and sepsis remains controversial, with several regimens being recommended:

- change catheter to a different site every 3–7 days
- change the catheter over a guide wire every 3–7 days
- do not change unless there is a problem.

In practice, no advantage has been documented to favour any one of these policies and the most important factor in the protection against catheter-related sepsis is the adoption of a strict aseptic technique when placing and caring for catheters. Several adjunctive techniques have been used to minimize the incidence of catheter-related sepsis: bonding antibiotics and silver compounds to the catheters, sterile protective plastic sleeves, silver ion-impregnated collagen cuffs, etc. Especially in ICU patients, catheter-related sepsis may have potentially lethal consequences, such as subacute bacterial endocarditis and suppurative thrombophlebitis.

The algorithm for management of catheters in critically ill patients is shown in Figure 11.2.

Fungal infections

These are important in ICU patients for two reasons:

1. Disseminated fungal infections constitute a grave prognostic marker of critical illness.
2. Despite newer fungal agents, the mortality (usually from multisystem organ failure) remains high, ranging from 33% to 75%. By far the most common organism responsible in ICU patients is *Candida albicans*. Much less commonly, other fungi, e.g. *Aspergillus* spp., *Torulopsis glabrata*, *Mucor* spp., are responsible.

The risk factors for colonization and invasive infections are:

- multiple antibiotic usage over a prolonged period
- immune depression
- steroid therapy
- parenteral nutrition
- concomitant bacteraemia.

Specific bacterial infections of surgical importance

Gas gangrene

Gas gangrene is extremely rare in civilian practice (0.1/100 000 per annum). The essential factor required for spore germination and production of illness is reduced oxygen tension. This may result from severe contusion and laceration with necrotic tissue, devitalization of a wound by compression, impaired blood supply, or foreign bodies implanted in the depths of a punctured wound, e.g. shrapnel or pieces of clothing and soil, which induces tissue necrosis as a result of its high ionized calcium salts and silicic acid. The oxygen tension of a wound may be further lowered by coexisting infection with pyogenic organisms. Gas gangrene has been reported following injection of epinephrine (adrenaline) into the buttocks where the skin is often contaminated by clostridial spores from the patient's faeces.

Bacteriology and pathology

Gas gangrene is a mixed clostridial infection by saccharolytic (pathogenic) and proteolytic (saprophytic) organisms. The true pathogens are *Clostridium perfringens* (*C. welchii*), *Clostridium novyi* (*C. oedematiens*) and *Clostridium septicum*. Type A *C. perfringens* is the most important human pathogen. It produces α-toxin, which is a lecithinase and breaks down the phospholipid constituents of red cells with the production of haemolysis. Other exotoxins produced by some strains of *C. perfringens* include haemolysin, collagenase, hyaluronidase and deoxyribonuclease.

These exotoxins produce a cellulitis and a progressive myonecrosis. They ferment the muscle carbohydrate, with the production of lactic acid and gas (H_2, CO_2). The discharge from the wound is initially odourless. Spread of the necrosis occurs as a result of exotoxin release and ischaemia from pressure by gas and exudate within tight muscle compartments. The affected area becomes tense, oedematous and crepitant. At first, the dead muscle is odourless and brick red in colour. Progressive putrefaction by saprophytic clostridia (*Clostridium sporogenes, Clostridium histolyticum*)

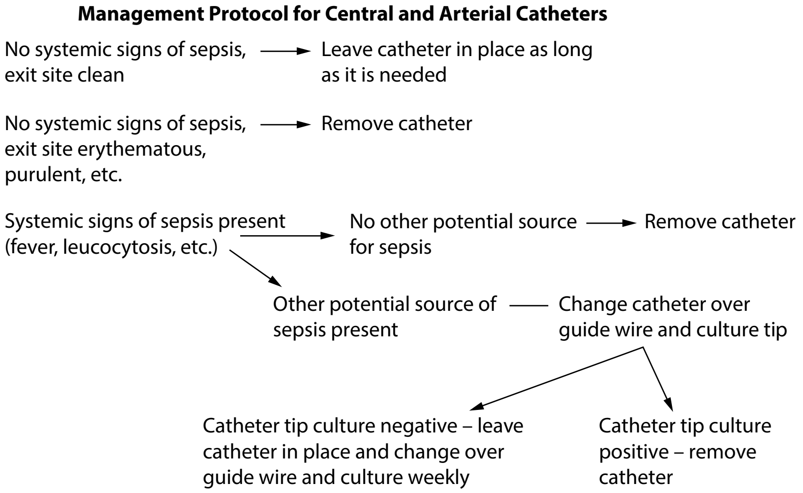

Management Protocol for Central and Arterial Catheters
No systemic signs of sepsis, exit site clean
Leave catheter in place as long as it is needed
No systemic signs of sepsis, exit site erythematous, purulent, etc.
Remove catheter
Systemic signs of sepsis present (fever, leucocytosis, etc.)
No other potential source for sepsis
Remove catheter
Other potential source of sepsis present
Change catheter over guide wire and culture tip
Catheter tip culture negative – leave catheter in place and change over guide wire and culture weekly
Catheter tip culture positive – remove catheter

Figure 11.2 Algorithm for the routine care of intravascular catheters.

of the dead muscle completes the pathological process with the production of the characteristic fishy odour and the greenish-black appearance of the established disease.

The profound toxaemia is due to the circulating exotoxins and results in shock, haemolytic anaemia, renal failure and jaundice. The organisms themselves do not invade the bloodstream except as an agonal event, and this accounts for the foamy liver and gas bubbles found in other organs at necropsy.

Clinical features

The majority of gas gangrene infections are exogenous and result from contamination of large wounds as occurs in agriculture tractor injuries, severe comminuted compound fractures sustained in road traffic accidents and battle casualties. An appreciable number of cases in civilian practice are, however, endogenous in origin from contamination by bowel organisms. In the West, endogenous gas gangrene is most commonly encountered following amputation for peripheral vascular disease. Risk factors in this group include diabetes and incontinence. Other instances of gas gangrene may result from criminal abortion and infections following intestinal and, less commonly, biliary operations.

The incubation period between the initiating incident and the onset of the clostridial infection varies from 1 day to 4 weeks. Its duration carries an inverse relationship with the severity of the illness and the mortality. The demonstration of gas (Figure 11.3) is not essential for the diagnosis as non-gas-forming clostridial infections of wounds are well documented. Indeed, the most useful clinical classification is into gas-forming and non-gas-forming clostridial infections. *The most important factor that determines whether the infection remains localized and non-crepitant, or becomes invasive with severe toxaemia and gas formation is the presence of dead muscle.*

Non-gas-forming infections

The disease is mild and, apart from pyrexia, there is minimal toxicity. The wound is oedematous and erythematous, and may develop a brownish discoloration. Crepitus is absent; pain and tenderness are not severe and the mortality due to the clostridial infection is negligible.

Gas-forming infections

The incubation period is usually less than 3 days and the onset acute. The condition declares itself by severe pain in the region of the wound and rapid development of toxaemia, drowsiness, fever and tachycardia. The affected area becomes swollen, tense, oedematous and extremely tender. The discharge may be serous or blood stained. It is variable in amount and initially odourless but subsequently becomes sweet to foul smelling. Gas is detected by crepitus and radiological examination. The overlying skin goes through a series of changes from intense white to inflammatory erythema with ecchymosis, bullae formation and frank greenish-black gangrene (Figure 11.4). In severe cases, jaundice, haemolysis and renal failure develop and contribute to the death of the patient. The overall mortality is 40%, but the mortality due to overwhelming clostridial infection is 10–15%.

Treatment

The treatment consists of general resuscitative measures for shock and specific therapy, i.e. antibiotics, antitoxin, surgical treatment and, when available, hyperbaric oxygen.

Antibiotic therapy

This is used both in the prophylaxis and in the treatment of established disease. For prophylaxis, the antibiotic therapy must be started immediately before the operation, e.g. amputation for peripheral vascular disease or soon after the injury, and should be continued until the healing is complete if the risk is high. The benefit of antibiotic therapy in established disease remains doubtful, largely because of poor antibiotic penetration into ischaemic tissue. Nonetheless, it constitutes part of the orthodox therapy. Antibiotic treatment should provide wide coverage to include Gram-positive

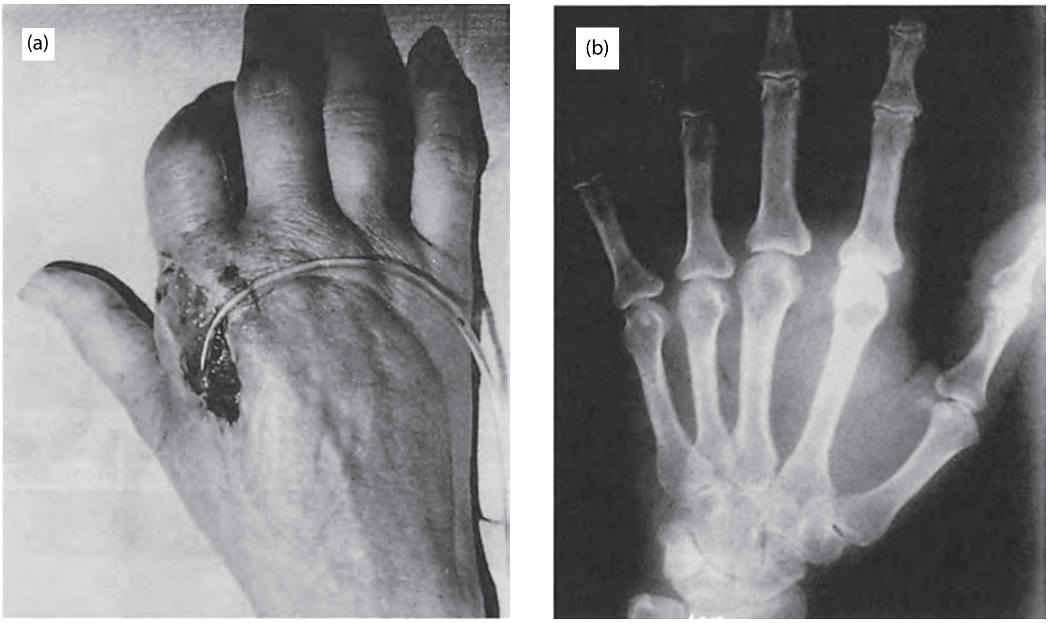


Figure 11.3 (a) Gas gangrene of the hand; (b) radiograph of the same hand showing gas in the soft tissues.

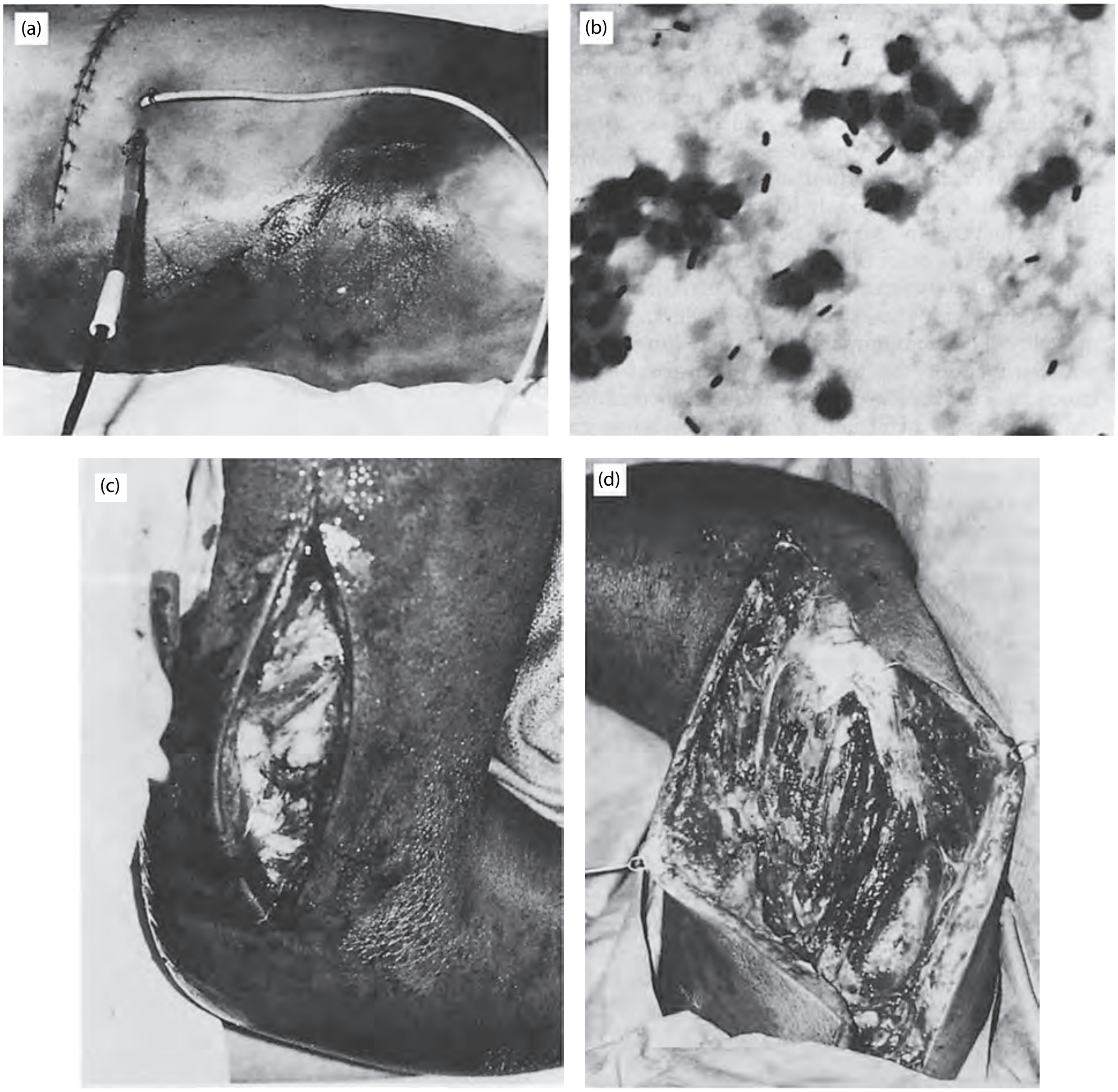


Figure 11.4 Gas gangrene: (a) abdominal wall after common bile duct exploration for ductal calculi; (b) microscopy of the needle aspirate from the same patient showing clostridial organisms; (c) gas gangrene of the upper arm: note bulging of oedematous tissues after incision of the skin; (d) same patient: extensive necrosis of upper arm muscles.

(penicillin or cephalosporin), Gram-negative (aminoglycoside, third-generation cephalosporin or ciprofloxacin) and anaerobic bacteria (clindamycin or metronidazole). In addition, vancomycin or linezolid should be considered in those at risk for MRSA. Tetracycline and erythromycin exhibit moderate activity against most clostridial species but rapid development of drug resistance is a problem. The treatment is continued for a minimum of 7 days and the antibiotic regimen may need to be changed, depending on the bacteriology and sensitivity tests.

Surgical treatment

Surgical treatment must be carried out immediately after resuscitation and commencement of antibiotic therapy and is delayed only if facilities for hyperbaric oxygen therapy are available (see below), when operative intervention is postponed until completion of the first hyperbaric treatment. The aim of surgical treatment is the excision of all necrotic tissue regardless of anatomical defects thus produced. Pus is evacuated and the completely debrided wound is irrigated with hydrogen peroxide solution. *The complete excision of all dead and infected tissue at the first operation is crucial to the survival of the patient.* In limb infections, this may necessitate amputation. No attempt is made to provide skin cover and the wound is packed with gauze soaked in isotonic saline solution. The patient is returned to the operating theatre 24–48 hours later for a dressing change under general anaesthesia. Any residual necrotic areas are excised down to bleeding tissues, after which the wound is dressed as before. Reconstructive surgery and skin grafting are delayed until the infection has been totally eradicated.

Hyperbaric oxygen therapy benefits patients with pure clostridial infections, and may result in rapid improvement in the clinical condition and in limb salvage. Hyperbaric oxygen therapy is started soon after the initial resuscitation and before surgical intervention. It consists of repeated treatments of 1.5–2 hours at a pressure of 250 kPa (2.5 atm).

Mortality

The reported mortality from traumatic gas gangrene is greater than 25% but that associated with non-traumatic gas gangrene caused by *C. septicum* is much higher, with reported mortality rates varying from 67% to 100%.

Clostridial enterocolitis and pseudomembranous colitis

This results from the ingestion of improperly cooked food contaminated by *C. perfringens*. The disease is usually self-limiting and causes severe colicky abdominal pain and diarrhoea. The organism is present in the stool in high counts. Occasionally, the condition is more severe and leads to widespread necrosis of the bowel (primarily of the small intestine) and is then referred to as enteritis necrotica. In addition to severe abdominal pain, vomiting and diarrhoea, the patient exhibits signs of peritonitis with profound toxaemia and shock. The condition carries a very high mortality.

Enterocolitis caused by *C. perfringens* has to be distinguished from PMC, which develops as a complication of antibiotic therapy in about 1 out of every 15 000 people treated with an antibiotic and is caused by *C. difficile*. The incidence of PMC is rising in many countries and approximately 30 000 people die each year from PMC in the USA. The risk factors for PMC include:

- old age
- AIDS
- suppression of gastric acid secretion
- antibiotic use
- chemotherapy
- Crohn disease
- enemas
- previous PMC infection
- prolonged hospitalization
- recent surgery
- tube feedings
- ulcerative colitis.

The symptoms of PMC colitis include diarrhoea that contains green or yellow mucus and blood in the stool, tenesmus, anorexia and nausea, cramping abdominal pain, vomiting, fever and rigors and weight loss. When suspected, stool cultures and stool test for *C. difficile* are mandatory in addition to other tests such as colonoscopy. Treatment is with rehydration and antibiotic therapy with any of the following depending on sensitivity: fidaxomicin, metronidazole, vancomycin or rifaximin (see Chapter 30).

Infective non-clostridial gangrene

Various clinical syndromes have been described as infective non-clostridial gangrene. The most common causative organisms are anaerobic streptococci, but necrotizing infections with *E. coli* and *Bacteroides* are well documented. The most common member of the *Bacteroides* species responsible for infective gangrene is *Fusiformis fusiformis,* which is often accompanied by *Borrelia vincentii,* although it is doubtful whether the latter plays any part in the disease process.

These gangrenous conditions usually arise against a background of debility, atherosclerosis and diabetes mellitus. The causative anaerobes often act in association with *S. pyogenes*, staphylococci and coliform bacilli. However, in some well-documented cases, a causative organism cannot be identified. In some cases a precipitating factor, e.g. trauma, operation or viral infection, initiates the condition; in others, particularly in diabetic patients, the morbid process arises spontaneously. A commonly favoured pathological classification is into:

- cutaneous gangrene (progressive bacterial gangrene)
- subcutaneous gangrene (necrotizing fasciitis)
- subfascial infective gangrene.

In cutaneous gangrene, the necrosis is limited to the skin only and systemic signs are usually minimal, although the disease may extend to the deeper tissues. In subcutaneous gangrene, the necrotic process primarily involves the subcutaneous fat and/or the deep fascia, usually sparing the underlying muscle layer. Necrosis of the skin is secondary to the development of thrombosis of the perforating vessels as they course through the necrotic infected deeper layers. Subcutaneous gangrene is a serious, rapidly spreading disease, which is accompanied by toxaemia and may prove fatal. Subfascial infections consist of myositis and myonecrosis.

The above classification, although useful in outlining the pathology and prognosis in the majority of these infections, does not cover the entire spectrum of these infections, some of which cannot be readily placed in either of these categories. In addition, a cutaneous infection may spread to involve the subcutaneous and subfascial compartments. Thus, a more comprehensive clinical classification of infective non-clostridial gangrene is:

- cutaneous gangrene (Meleney's progressive bacterial gangrene)
- subcutaneous gangrene: necrotizing fasciitis and Meleney's undermining ulcer
- infected vascular gangrene
- cancrum oris and noma vulva: protein calorie malnutrition
- streptococcal myositis
- human bite infections.

Cutaneous gangrene (Meleney's progressive bacterial gangrene)

This is usually a synergistic infection with two microbial species that differ in their requirements for oxygen. In classic cases, the micro-organisms involved are microaerophilic streptococci and aerobic staphylococci, although *Bacteroides* and a variety of Gram-negative aerobic bacteria can be involved. Bacterial inoculation most frequently follows minor trauma, or a postoperative consequence of a surgical drain or mass closure with deep tension sutures. It often follows drainage of abscesses, particularly in patients with diabetes mellitus and severe atherosclerosis. In most instances, 1–2 weeks elapse before the onset of clinical manifestations. At that time, the skin surrounding the wound becomes red, oedematous and very tender. As the cutaneous infection extends outwards, the central area becomes necrotic, turning to eschar that ulcerates (Figure 11.5). A thin, foul-smelling discharge simulating dishwater emanates from the ulcer. At this stage, the patient becomes systemically toxic. Broad-spectrum antibiotics and surgical excision are necessary for survival. The most favoured first-line antibiotic combination is benzyl penicillin, metronidazole and gentamicin. At operation, the infected necrosis is found to be limited to the skin and subcutaneous fat, but the fascia and underlying muscle are spared except for in advanced cases. Following complete excision, the wound is left open and packed with gauze soaked in a bactericidal agent. Debridement may have to be repeated. Secondary wound closure or skin cover by grafting is only attempted after the infection has been completely eradicated and healthy granulation tissue has formed.

Subcutaneous gangrene

The most important example of this is necrotizing fasciitis, also known as haemolytic streptococcal gangrene, hospital gangrene and gangrenous erysipelas. It includes conditions such as perineal phlegmon (Figure 11.6) and Fournier's scrotal gangrene (Figure 11.7). It is caused by haemolytic streptococci and, less commonly, haemolytic staphylococci. Various other organisms have been identified in some of these infections, including coliforms, *Bacteroides*, diphtheroids and *Pseudomonas.* Most commonly, the condition arises following surgery or trauma. Spontaneous cases have been described, although, in some of these patients, the preceding trauma may have been so slight as to be ignored by the patient. The most commonly affected sites are the extremities, followed by the lower trunk, including the external genitalia and the perineum. The exact mechanism for the subcutaneous necrosis is unknown but appears to be related to the binding of the mucopeptide fraction of the bacterial cell wall with dermal collagen. The necrosis does not involve the muscle layer and skin involvement is secondary to thrombosis of the perforating vessels coursing through the infected necrotic area.

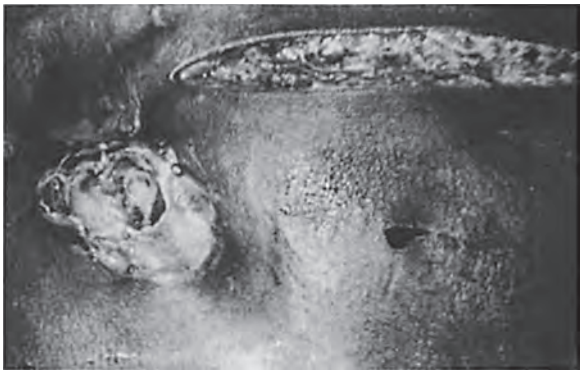


Figure 11.5 Meleney's progressive bacterial gangrene. Culture revealed *Staphylococcus aureus* and a microaerophilic *Streptococcus.*

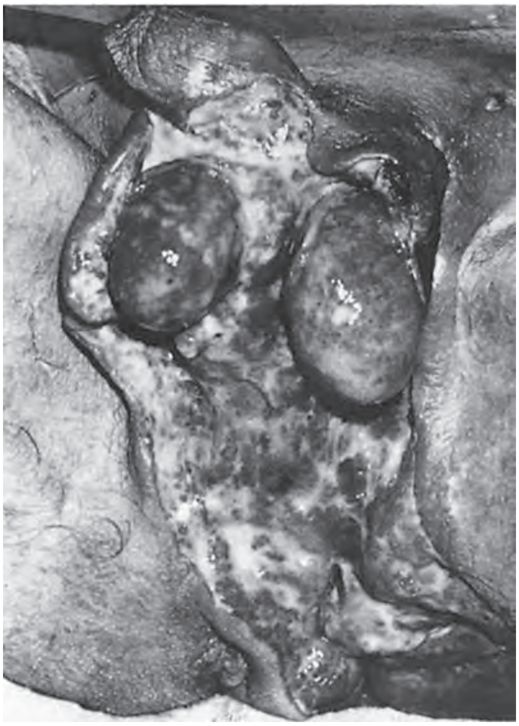


Figure 11.6 Extensive perineal phlegmon involving the skin of the scrotum and penile shaft.

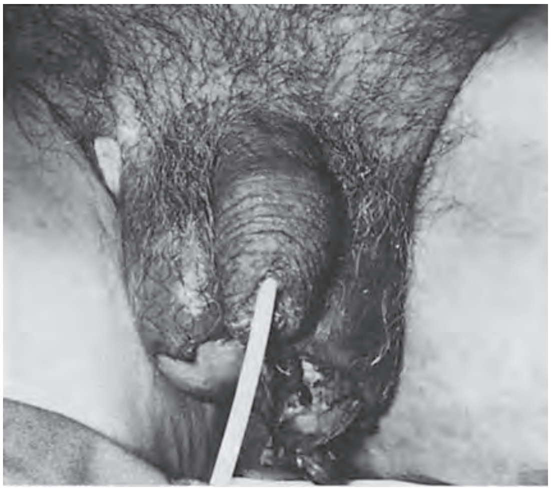


Figure 11.7 Fournier's gangrene with necrosis of scrotal skin.

The disease is always serious and carries a definite mortality. The affected part is initially very painful but then becomes numb owing to the involvement of sensory nerve fibres. The process spreads rapidly through the subcutaneous fatty/fascial plane with reddish discoloration, inflammatory oedema (sometimes bullous), necrosis and eventual sloughing of the overlying skin (Figure 11.8). Systemic manifestations are always present and the toxaemia may be severe with pyrexia, tachycardia and shock. Blood cultures should always be taken in these patients. Treatment includes resuscitation with crystalloids and blood, antibiotics using a triple regimen (benzyl penicillin, metronidazole and aminoglycoside), and early wide surgical excision and drainage with delayed skin cover.

Meleney's chronic undermining ulcer results from infection by a microaerophilic *Streptococcus* and usually develops after surgery on the intestinal and genital tracts. The infection and necrosis start in the subcutaneous tissues (Figure 11.9) but the disease progressively affects the deeper tissues to involve the pelvis. Treatment is with a combination of benzyl penicillin and metronidazole, together with surgical excision and delayed skin cover.

Infected vascular gangrene

A significant proportion of these necrotic infections occur in diabetic patients. Infected vascular gangrene has a gradual onset. The affected part (usually the foot) becomes painful, swollen, black and foul smelling. Radiographs show extensive gas formation in the tissues of the involved foot. The infection is a mixed one with faecal organisms, most commonly *B. fragilis* and *Peptostreptococcus.* The severity of the toxaemia is variable.

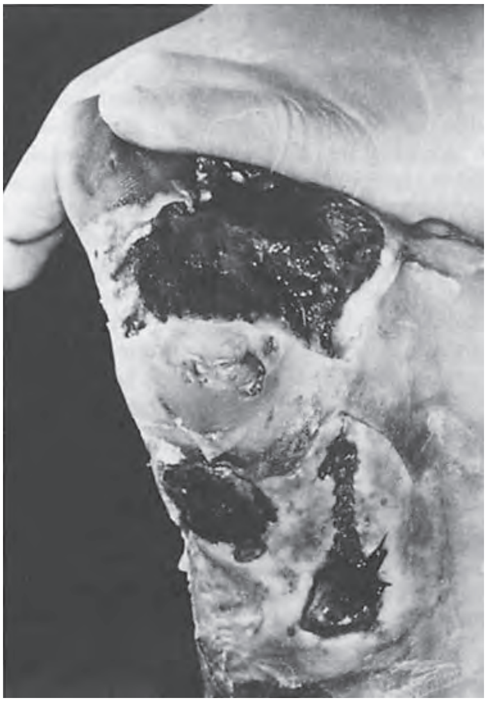


Figure 11.8 Necrotizing fasciitis of the sole of the foot.

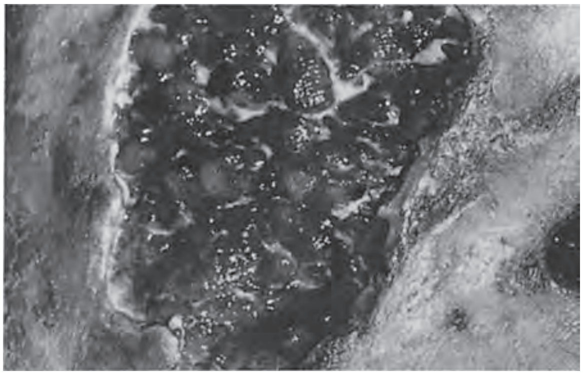


Figure 11.9 Meleney's chronic undermining ulcer in the right iliac fossa in a renal transplant patient.

Streptococcal myositis

This represents a unique infection that can usually be distinguished from clostridial infections by the presence of pronounced skin reaction: cyanotic discoloration, blebs containing foul-smelling fluid teeming with Gram-positive cocci, and islands of frankly gangrenous skin. After instituting supportive care and antibiotics, treatment is with a generous incision, exploration and drainage of all the infected muscle groups, aptly termed the 'fillet' procedure. Only occasionally is myonecrosis present and this requires excision. The wound is kept open and packed with saline gauze. Prognosis is surprisingly good if the infection is managed before the onset of myonecrosis.

Cancrum oris and noma vulvae

These are instances of mucocutaneous gangrene affecting the mouth (cancrum oris) or the vulva (noma vulvae). Both arise against a background of malnutrition in children and are usually preceded by an infectious illness, such as measles. The infection is often a mixed one, but the protagonists are either anaerobic streptococci or members of the *Bacteroides* species. The disease results in slow but relentless necrosis of the perioral or vulval tissues, and in cancrum oris death may result from inhalation pneumonia. Correction of the underlying malnutrition is essential, in addition to antibiotic therapy. Skin cover and plastic reconstruction are delayed until the infection has been cleared and the patient's nutritional status has improved.

Human bite infections

The anaerobic infections caused by human bites can be particularly virulent and cause marked tissue destruction. The infection is usually a mixed one, the causative organisms being a combination of two or more of the following: *Bacteroides melaninogenicus*, *Fusobacterium* spp., anaerobic cocci and spirochaetes. The first-line antibiotic is penicillin, which is administered in high doses. The importance of tetanus prophylaxis in these patients must not be forgotten.

Tetanus

In the vast majority of cases, tetanus is an exogenous infection, although rare instances of endogenous infections are documented after septic abortions and operations on the gastrointestinal tract. The conditions governing the germination of spores of *Clostridium tetani* are identical to those of gas gangrene and necessitate local hypoxia. The most common portal of entry worldwide is the umbilical stump following the application of dung to this region in newborn babies that is practised in some developing countries. Elsewhere, the lower limbs constitute the most common portal of entry. Usually, the wound is a minor one but it is always penetrating in nature. *In some 25% of cases in the West, the portal of entry is not evident at the time of diagnosis.*

Other sources of infection include piercing of the ear lobes; tattooing; burns; parenteral injections including vaccination; skin lesions, especially leg ulcers; nasal foreign bodies; and ear infections. The disease is well recognized in drug addicts.

C. tetani (Figure 11.10) produces two exotoxins. The most important is a neurotoxin called tetanospasmin that is responsible for the disease. A haemolytic toxin called tetanolysin acts on the peripheral neuromuscular junctions but does not play a significant role in the disease. Tetanospasmin reaches the central nervous system (CNS) along the axons of motor trunks, probably in the tissue spaces between the nerve fibres, and acts by blocking the inhibitory impulses at the motor synapses. This results in two forms of contractions of striated muscles: tonic (spasm), characteristic of the early disease, and clonic (convulsions), indicating severe established disease.

The overall mortality of tetanus is 10–15%. Adverse factors include extremes of age, short incubation period, type of injury and severity of the illness. Thus, whereas mild to moderate tetanus carries a small mortality, death from the disease occurs in 30–40% of patients with severe tetanus. Death usually results from asphyxia from involvement of the muscles of respiration and from cardiovascular complications resulting from sympathetic overactivity.

Clinical features

Neonatal tetanus is by far the most common type in developing countries, where it accounts for 70% of cases. It is best considered separately from the disease in children and adults.

Neonatal tetanus

The clinical picture is characteristic and the disease usually becomes manifest on the eighth day, hence the popular name 'eighth-day disease'. It starts with failure to suckle on the third day. This is followed quickly by spasm of the facial muscles (risus sardonicus) and masseter (lockjaw), and progression to generalized clonic spasms with flexion of the arms, clenched fists, extension of the lower limbs and plantarflexion of the toes.

Children and adults

The majority of patients (95%) who develop tetanus have not been previously immunized. The incubation period ranges from 4 to 10 days, and the shorter the interval the more severe the disease. The progress of the disease (the time from the first symptom to the onset of tetanus) varies according to the severity of the disease but the full-blown picture is reached by the third day in 70% of cases. The condition declares itself by stiffness, twitching and cramps limited to the same spinal segment as the area of infection (local tetanus). Other early symptoms and signs include muscle pains, headaches, irritability, restlessness, constipation, sweating and tachycardia. This is followed by the development of spasm of the masseter muscles, facial musculature and muscles of deglutition (dysphagia). In the full-blown picture, there are generalized clonic convulsions that may be triggered by mild external stimuli (sound, movement of personnel, etc.). As the extensor muscles are more powerful than the flexor muscles, the patient classically assumes a position of opisthotonos. It is characteristic of tetanus that the muscles do not relax between convulsive attacks, and this distinguishes tetanic convulsions from those caused by strychnine poisoning.

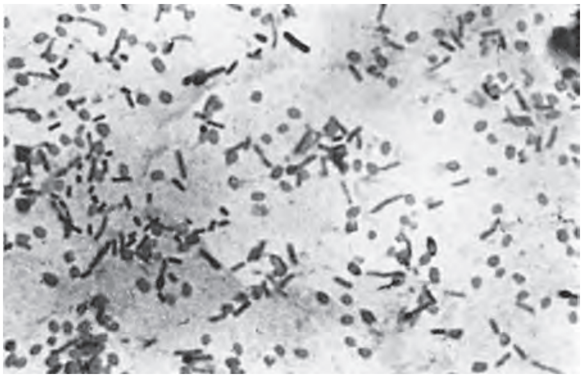


Figure 11.10 *Clostridium tetani* organisms.

Tetanus is classified into:

- mild: no dysphagia or respiratory distress
- moderate: presence of dysphagia and respiratory distress
- severe: gross spasticity and major spasms.

Treatment

Prophylaxis

The best method of prevention is by active immunization with tetanus toxoid (a formolized preparation of the exotoxin adsorbed on aluminium hydroxide or phosphate). Three injections are administered, with the second injection 6 weeks after the first, and the third injection 6–12 months later. A booster injection is given at 10 year intervals and at times of wounding. Booster toxoid injections are not necessary if a patient sustains a wound within 5 years of completion of an active immunization course or booster dose. However, if the period since the last toxoid injection exceeds 5 years, but is less than 10 years, a booster dose should be administered. In the absence of a history of active immunization, or if the period since the last toxoid injection exceeds 10 years, passive immunization with human tetanus immunoglobulin [tetanus immune globulin (TIG)] in a dose of 250 units intramuscularly is indicated. This single dose provides immunity for about 4 weeks. If the wound has not healed by this time, a second dose is administered. Active immunization with toxoid should be started at the same time (using a different limb) in those not previously immunized, or a single booster dose is administered to those patients who had allowed their active immunization to lapse for more than 10 years. The use of equine antitetanic serum is no longer practised because of the risk of anaphylactic reactions. Active immunization with tetanus toxoid of the pregnant female will protect the infant from neonatal tetanus.

Treatment of established disease

Specific measures include surgical attention to the wound if present with excision and open packing with antibiotic or hydrogen peroxide-soaked gauze. Benzyl penicillin is administered in a dose of 2 MU every 4 hours for 7 days, and TIG is given in a dose of 2000–4000 U intramuscularly. Metronidazole (500 mg orally every 6 hours or 1.0 g rectally 8 hourly) may be used instead of penicillin and is reported to be more effective. In neonatal tetanus, good results have been obtained with intrathecal TIG and prednisolone. Doxycycline is another effective antibiotic. Supportive measures include intravenous fluids, control of pain and sedation. Diazepam alone, compared with other sedatives such as phenobarbitone, appears to more effective in treating tetanus, and adding other drugs to diazepam does not appear beneficial and may be harmful.

For mild tetanus, the patient only requires sedation with diazepam. Tracheostomy and sedation are necessary for patients with moderate tetanus. In severe disease, the patient has to be treated with neuromuscular blockade and positive-pressure ventilation. The very severe cases with sympathetic overactivity and cardiovascular complications (tachycardia, labile hypertension, vasoconstriction, myocardial instability) require general anaesthesia, mechanical ventilation and adrenergic blockade.

Chronic bacterial infections

These are produced by the species of the order Actinomycetales, so named because the bacterial cells may branch to form hyphae. With some species, e.g. *Actinomyces israelii,* this branching tendency is marked and simulates the appearance of a fungal growth (*actino*, radial; *myces*, fungus). The composition of the order Actinomycetales is shown in Table 11.1.

Streptomycetaceae are not pathogenic to humans. The genus *Streptomyces* is, however, of medical importance as it contains many species valuable for the production of antibiotics (streptomycin, tetramycin, etc.). Most species of *Nocardia* are harmless and live on decaying organic matter in the soil. A few species, e.g. *Nocardia asteroides*, can cause infection of the lung that is often fatal and may be mistaken for tuberculosis. Other species (*Nocardia madurae, Nocardia brasiliensis*) are known to produce a chronic disease of the hands or feet similar to actinomycosis.

Tuberculosis

In conjunction with all of the mycobacteria, the tubercle bacilli are non-sporing, immobile, aerobic and Gram positive. Differential staining methods are used to identify the mycobacteria since, after heat staining with carbol fuchsin, these bacteria, with the exception of *Nocardia* spp., are unique in resisting decolouration after treatment with strong acids and alcohol, i.e. are acid fast.

There are several strains of non-pathogenic saprophytes in the soil and on plants (e.g. *Mycobacterium phlei*) and on the human skin (*Mycobacterium smegmatis*). The pathogenic status of several other species is variable but some are often found in tuberculosis-like disorders or in association with established tuberculosis, especially in immune-suppressed individuals. They are referred to variously as atypical, environmental, anonymous or MOTT (mycobacteria other than typical tubercle) to differentiate them from the established pathogens, i.e. *Mycobacterium tuberculosis* and *Mycobacterium bovis.* The precise differentiation of the various mycobacteria is important. Culture is usually performed on Lowenstein–Jensen medium and requires 6 weeks, although faster growth can be obtained in Dubois medium. No single test is said to be reliable in the identification of *M. tuberculosis* and *M. bovis*, although tests for virulence in guinea-pigs (*M. tuberculosis*) and rabbits (*M. bovis*) are among the most reliable of the diagnostic tests. *Mycobacterium tuberculosis* is the only species known to produce niacin on culture.

Table 11.1 The Actinomycetales

Order	Family	Genus	Pathogenic species
Actinomycetales	Mycobacteriaceae	*Mycobacterium*	*M. tuberculosis*
			M. leprae
	Actinomycetaceae	*Actinomyces*	*A. israelii*
		Nocardia	*N. asteroides*
			N. madurae
	Streptomycetaceae	*Streptomyces*	–

Following the eradication of tuberculous herds and the introduction of pasteurized milk, bovine tuberculosis is rarely encountered in the West, and most of the reported cases are pulmonary infections. The vast majority of cases in these countries are caused by *M. tuberculosis*, usually as a result of inhalation of organisms present in fresh droplets or dust contaminated with dried sputum by a patient with open pulmonary tuberculosis. Both bovine and human infections are still common in economically deprived areas, such as parts of the African continent, Latin America and India. The global increase in tuberculosis that occurred during the 1980s and 1990s was associated with a re-emergence of resistance to antituberculous drugs (isoniazid and multidrug).

Pathology

It has been estimated that 90% of all tuberculous infections involve the lungs, but the infection may affect practically any organ or tissue. The most commonly recognized extrapulmonary infections include tuberculosis of the skin, lymph nodes, bones and joints, genitourinary tract, abdomen and intestines, and CNS.

Tuberculous lesions assume one of two forms:

- proliferative: the tubercle follicle
- exudative.

The most common proliferative lesion that is usually encountered in the lungs and solid organs is the tubercle follicle. This consists of an area of coagulative necrosis (caseation) due primarily to hypersensitivity to the tuberculoprotein, surrounded by epithelioid and Langerhans giant cells (both derived from macrophages), and an outer zone of small, round cells consisting mainly of lymphocytes and fibroblasts.

The exudative form of tuberculosis is typically encountered in infections of the serous cavities, e.g. tuberculous pleurisy/peritonitis, and epithelial surfaces (sterile pyuria in renal tuberculosis). It results in the formation of a cellular exudate rich in fibrin, together with a dense infiltration of the tissues with lymphocytes.

Childhood tuberculosis is characterized by marked involvement of the regional lymph nodes, as exemplified by the primary complex (Ghon focus at the periphery of the lung mid-zone and hilar lymphadenopathy) and tabes mesenterica, where a small focus in the intestine is associated with marked enlargement of the mesenteric lymph nodes, which at times rupture causing tuberculous peritonitis. By contrast, in the adult, lymph node enlargement is not marked and the disease either heals by fibrosis or extends locally by caseation, liquefaction and cavitation, with little tendency to bloodstream dissemination. This altered tissue response in the adult appears to be the result of tissue maturation.

Softening and liquefaction of the caseous material underlies the development of tuberculous 'cold' abscesses. The liquefied debris may track along fascial planes, as in the psoas abscess originating from spinal tuberculosis (Figure 11.11), or point to the surface with the eventual formation of tuberculous sinuses, e.g. collar-stud abscess in the neck from tuberculous lymphadenitis and scrotal sinuses from tuberculous epididymitis.

Involvement of a pulmonary vein by a tuberculous focus in the lung may lead to bloodstream dissemination and miliary tuberculosis, especially if the resistance is lowered by poor nutrition, debility, old age or immune deficiency from any cause. If the bloodstream inoculum is small, the bacilli may be destroyed by the cells of the monocyte–macrophage system. Failing this, they may either produce metastatic disease immediately or remain quiescent with reactivation some years later. These lesions are referred to as local metastatic tuberculosis or secondary tuberculosis and account for the majority of tuberculous infections encountered in surgical practice.

Clinical features

The general symptoms of active tuberculous infections include malaise, asthenia, weight loss, mild fever and night sweats. The symptomatology is, however, extremely varied and the disease may simulate many other disorders. The specific symptoms relate to the organ involved in the disease. Tuberculosis is a disease of malnutrition and overcrowding. Other predisposing factors include poor general health, chronic disease, silicosis and diabetes. Certain ethnic groups, such as the Australian Aborigines, black Africans and native Americans, are particularly susceptible.

Allergy (hypersensitivity) and acquired immunity

Tuberculous infection, subclinical or otherwise, results in the development of a cell-mediated allergy (delayed hypersensitivity) to tuberculoprotein that causes caseation and an accelerated macrophage response. This hypersensitivity, which indicates present or past infection, can be determined by the tuberculin skin reaction, which consists of the intracutaneous injection of purified protein derivative that is derived from and has replaced Koch's old tuberculin. The delayed hypersensitivity, although closely related, is not the mechanism of the acquired immunity to tuberculosis. This immunity, which is only partial, is cell mediated by sensitized lymphocytes. Active immunization with bacille Calmette–Guérin (BCG) is generally recommended for tuberculin-negative individuals. Complications of BCG vaccination are rare and include local abscess formation, regional lymphadenitis and, rarely, systemic infection with progressive pulmonary disease. The last has been reported following immunotherapy with BCG in patients with malignant disease.

Treatment

Effective modern chemotherapy is followed by a high cure rate without the necessity for long-term follow-up unless disease is caused by resistant organisms. It has eliminated the need for sanatorium management. The drugs available include streptomycin, para-aminosalicylic acid, isoniazid, rifampicin, pyrazinamide and ethambutol. *Mycobacterium bovis* is intrinsically resistant to pyrazinamide. The resistance to isoniazid in the UK now averages 5.5% and multidrug resistance 1.35%. Both

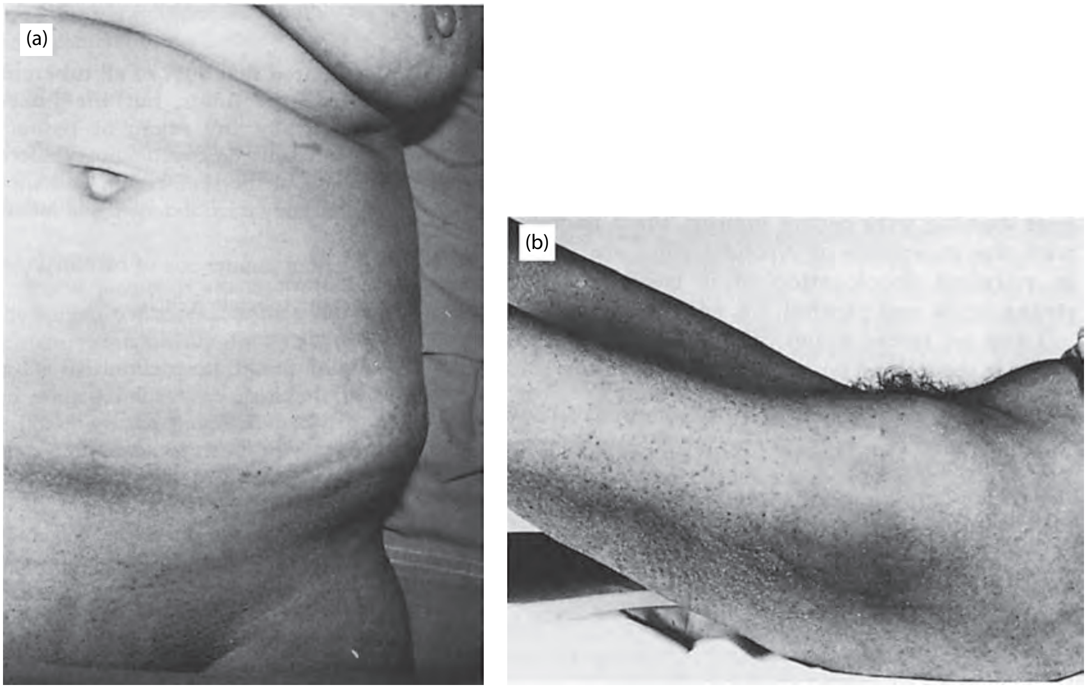


Figure 11.11 Tuberculous abscess: (a) left iliac fossa; (b) femoral triangle: infrainguinal extension of psoas abscess resulting from spinal disease.

isoniazid and multidrug resistance rates are higher in patients known to be infected with human immunodeficiency virus (HIV) (13.5% isoniazid and 6.1% multidrug).

Leprosy

It is estimated that 15 million people in the world have leprosy, which is common throughout most of Africa, southern Asia, the Far East, and South and Central America. There are about 400 active cases registered in the UK. However, no case of indigenously contracted leprosy has been reported in Britain.

Mycobacterium leprae, the causative organism, is a slender and acid-fast rod, occurring singly or in clusters (globi) in the reticuloendothelial (macrophage) cells. This slowly multiplying organism has been grown in the footpads of mice and the armadillo but not in an artificial culture medium. The infectivity of leprosy is a function of the concentration of leprosy bacilli in the body of the patient, and the chances of bacteria emerging and remaining viable and pathogenic to susceptible contacts. Leprosy is not a very contagious disease. It is frequently contracted in childhood or adolescence, revealing itself in symptoms and signs some years later. This silent period is commonly 2–5 years and often longer. Bacilli-laden nasal discharge is probably the main source of infection, but bacilliferous ulcerations, sweaty hairy skin and maternal milk may contain viable bacilli. Infection may be acquired by inhalation or through abrasions in the skin.

Pathology

The pathological lesions depend on the type and extent of the immune response. If cell-mediated immunity is strong, *tuberculoid* lesions form. The reaction consists of non-specific accumulation of giant cells, epithelioid cells, histiocytes and lymphocytes. Lymphocytic infiltrations are observed around and within nerve bundles. Bacilli are scanty *(paucibacillary disease)* and the lepromin test is positive.

When cell-mediated immunity is depressed, *lepromatous* leprosy (multibacillary disease) results. The whole dermis is replaced by highly bacilliferous tissue that invades the adnexa and eventually destroys superficial nerves, pigment-forming cells, sweat and sebaceous glands and hair follicles. The target tissues are the Schwann cells, endothelial cells and muscle cells. Bacilli may also be found in the liver, bone marrow, spleen, kidneys and lungs. Acute vasculitis caused by immune complexes may give rise to erythema nodosum leprosum, or other manifestations such as iritis, neuritis, orchitis, lymphadenitis and myositis. Longstanding lepromatous leprosy can result in chronic nephritis and amyloidosis.

Clinical features

Leprosy encompasses a whole spectrum of disease between the two main types, lepromatous and tuberculoid. The earlier manifestation, which is qualified as indeterminate, consists of a small hypopigmented macule, 2–5 cm, appearing anywhere in the body and often healing spontaneously. Determinate lesions that progress to clinical disease may arise out of indeterminate ones or *de novo*. Their subsequent features will depend on whether the disease is predominantly tuberculoid or lepromatous in nature.

Tuberculoid leprosy

In this type (Figure 11.12), there are a few or solitary skin lesions measuring 2–5 cm in diameter, often with a raised edge. The lesions are dry, hairless and anaesthetic. They are hypopigmented in dark-skinned and coppery in white-skinned individuals. Local or distant cutaneous nerve thickening occurs. The most common nerves involved are the ulnar nerve at the elbow, the median nerve at the wrist, the common peroneal in the popliteal fossa, the posterior tibial around the medial malleolus and the greater auricular nerve over the sternomastoid muscle. Sensory, motor and autonomic nerve trunks are affected. Nerve damage occurs early in tuberculoid leprosy.

Lepromatous leprosy

This presents as a widespread symmetrical macular rash, slightly hypopigmented or erythematous, affecting the face, extensor surface of the limbs and the upper trunk. The midline of the back, the axillae, groin and scalp are usually spared. There is congestion and discharge from the nose, the mucosa of which is thickened and yellow. Iritis may occur.

As the disease progresses, the skin, especially of the face, becomes thickened and nodular with thinning or loss of the eyebrows (Figure 11.13). Symmetrical enlargement of the peripheral nerves occurs, with widespread peripheral anaesthesia and muscular weakness. Painless neuropathic ulceration, absorption of the extremities of the digits, wrist drop, foot drop and claw hand occur (Figure 11.14). Nerve damage is observed late in lepromatous leprosy. Although acute exacerbations are found in all kinds of leprosy, they are most serious in the lepromatous type. An acute relapse is often heralded by an attack of erythema nodosum leprosum, and manifests widespread sensitivity phenomena in the skin, uveal tract, nerves, lymph nodes and joints.

Borderline leprosy

The presenting features are not typical of either tuberculoid or lepromatous leprosy. The disease may be arrested at this stage or progress in either direction. In borderline tuberculoid leprosy, the lesions are more numerous and varied than in the pure tuberculoid disease. The peripheral nerves are thickened and sensation is impaired. It is a common presentation in Africans. Asians and Europeans usually present with borderline lepromatous leprosy. Symmetry of the skin lesions is less

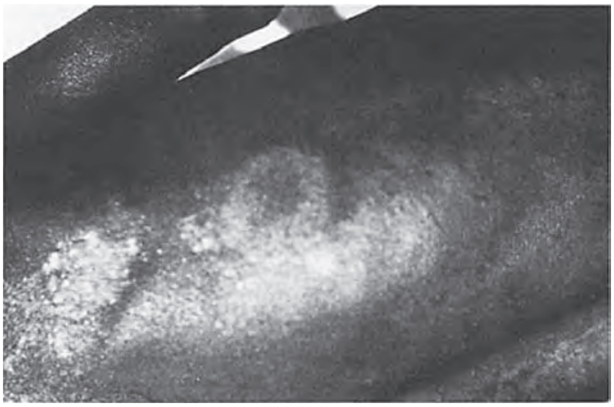


Figure 11.12 Tuberculoid leprosy.

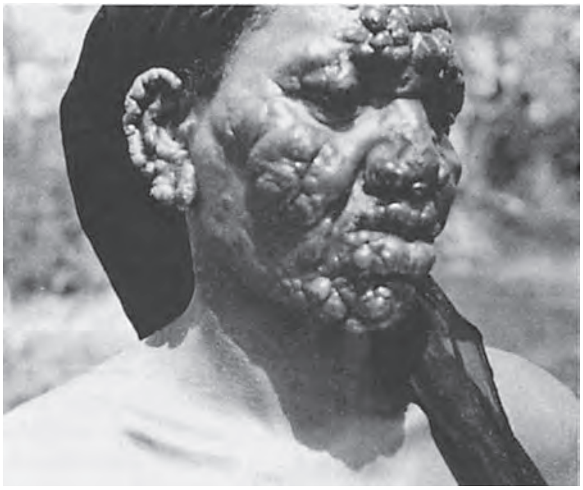


Figure 11.13 Lepromatous leprosy.

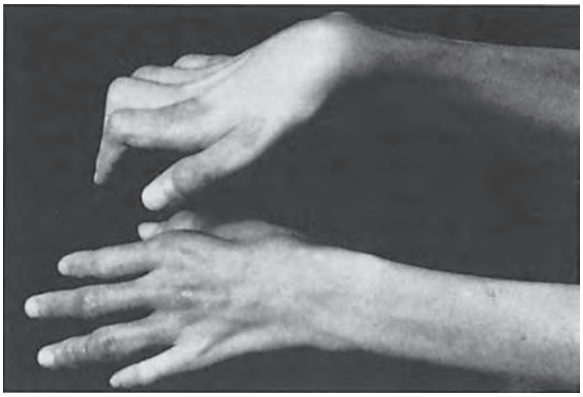


Figure 11.14 Claw hand in advanced lepromatous leprosy. With early treatment this lesion should never be seen.

constant, while the nodules are often discrete, red and fleshy. Nerve damage occurs early.

Diagnosis

Examination of dermal material obtained by the slit–scrape method reveals numerous bacilli in lepromatous and borderline lepromatous leprosy, while in paucibacillary disease (tuberculoid and borderline tuberculoid) bacteria are scanty and difficult to demonstrate. The dermal fluid is placed on a slide, dried and stained with a modified Ziehl–Neelsen method. Bacilli are also found in the nasal discharge. Histological examination of biopsies confirms the diagnosis.

Treatment

Multibacillary disease

The treatment of lepromatous and borderline lepromatous leprosy is by a combination of (1) rifampicin 600 mg once monthly, (2) dapsone 100 mg daily and (3) clofazimine 50 mg daily with an additional monthly dose of 300 mg. Where clofazimine is unacceptable because of its effects on skin colour, it is replaced with ethionamide 250 mg. This triple drug therapy is given for a minimum of 2 years, preferably until such time as the patient has achieved skin smear negativity. Acute exacerbations of multibacillary disease require hospitalization with complete physical and mental rest. If the symptoms are not relieved by aspirin 0.6 g three times a day, chlorpromazine is administered for 5 days. Patients with severe symptoms or those who develop erythema nodosum leprosum lesions require treatment with prednisolone or clofazimine or thalidomide.

Paucibacillary disease

Tuberculoid and borderline tuberculoid leprosy are treated with (1) rifampicin 600 mg once monthly and (2) dapsone 100 mg daily for 6 months.

Actinomycosis

This is a rare chronic infection caused by *A. israelii,* most commonly in the region of the lower jaw (cervicofacial). Actinomycosis occurs worldwide and is commoner in areas with low socioeconomic status and poor dental hygiene. The disease is characterized by the formation of loculated abscesses with marked induration and sinus formation (Figure 11.15).

Pathology

Actinomyces israelii occurs as a normal commensal in the human mouth. The organism is anaerobic and Gram positive but not acid fast. Although the precise conditions that result in the development of this endogenous infection are not known, the disease often follows trauma such as extraction of a carious tooth. The traumatic implantation of the organism in sufficient numbers appears necessary for the establishment of the disease, and cases following human bites or penetrating hand injuries resulting from violent contact with human teeth (punch actinomycosis) are well documented. The disease starts as an area of acute suppurative inflammation that persists as a chronic process with the formation of multiple loculated abscess cavities surrounded with dense fibrosis. Colonies of the organism occur in the pus as small greyish-yellow granules (sulphur granules), and consist of a densely felted mass of filaments surrounded by radially disposed club-shaped excrescences (Figure 11.16).

The disease spreads mainly by direct contact, with considerable destruction of tissue and multiple sinus formation. Blood-borne spread is important, as exemplified by the spread of ileocaecal actinomycosis via the portal vein to the liver, with the development of multiple intercommunicating liver abscesses (honeycomb liver). Pulmonary infection may also disseminate via the bloodstream to other organs such as bones, kidneys and

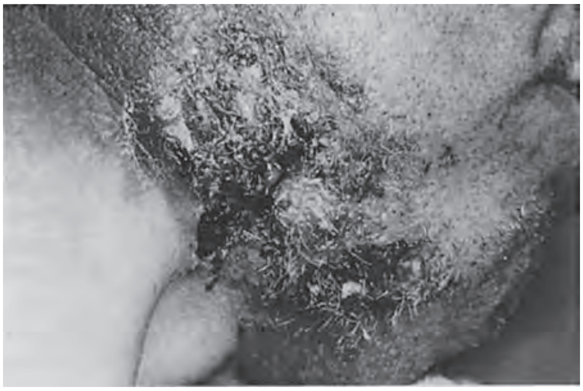


Figure 11.15 Cervicofacial actinomycosis.

CNS. In most instances, actinomycosis starts in the cervicofacial region (70%). In other instances, the primary infection occurs in the ileocaecal region (20%) or lungs (10%). For unknown reasons, actinomycosis is more common in men than in women (male–female ratio, 3:1), with the exception of pelvic actinomycosis.

The infection can affect all ages, but most cases are reported in young to middle-aged adults (aged 20–50 years). Nowadays, mortality is low because of the efficacy of antibiotic therapy. Preliminary diagnosis can be made by examining sulphur granules. The granules are squashed between two slides and stained with 1% methylene blue before microscopic inspection. Diagnosis is confirmed by culture (aerobic and anaerobic). A Gram-stained smear of the specimen may demonstrate the presence of beaded, branched, Gram-positive filamentous rods, suggesting the diagnosis. Cultures should be incubated for 48 hours or longer; the isolation and definitive identification of actinomycetes may require 2–3 weeks. Nucleic acid probes and polymerase chain reaction (PCR) methods have been developed for more rapid identification. Antimicrobial susceptibility testing is not needed as actinomycetes are invariably sensitive to penicillin, although the copathogens may not be.

Treatment

The organism has a wide spectrum of sensitivity to commonly used antibiotics such as penicillin and lincomycin. Prognosis is good following prolonged antibiotic treatment.

High-dose penicillin administered over a prolonged period (6 months to 1 year) is the established treatment. More recently, shorter courses of penicillin therapy (<6 months have been reported as sufficient especially in cervicofacial actinomycosis. However, the duration of penicillin therapy should be tailored to the individual patient based on clinical and radiographic response. The risk of actinomycetes developing penicillin resistance is minimal. Lack of a clinical response to penicillin is due to the presence of resistant companion bacteria, which may require modification of the antibiotic regimen, i.e. addition of an antibiotic active against these copathogens.

In patients who are allergic to penicillin, the following antibiotics may be used:

- *lincosamide*: lacks activity against some companion bacteria
- *amoxicillin/clavulanic acid (Augmentin)*: covers both pathogenic actinomycetes and companion bacteria, which frequently are resistant to penicillin
- *ceftriaxone*: covers both pathogenic actinomycetes and companion bacteria; used in cervicofacial and thoracic actinomycosis
- *imipenem/cilastatin*: covers both pathogenic actinomycetes and companion bacteria, which frequently are resistant to penicillin; used in abdominal and pelvic actinomycosis. See also Chapter 29.

Syphilis

This venereal disease is caused by *Treponema pallidum* and is transmitted by sexual intercourse. The most common portal of entry is the genital region, followed by the mouth or lips. Transmission by fomites is rare since the organisms are destroyed by rapid drying, but can occur. Placental transmission of *T. pallidum* is well documented with prenatal infection. In addition, infants may acquire extragenital infection during delivery from a mother with acute syphilis. Infants with congenital syphilis have lesions at birth or acquire them soon afterwards.

Pathology and clinical course

Following penetration of the skin or mucous membrane, the spirochaetes spread along the lymphatics and lymph nodes to reach the bloodstream within hours of exposure. The primary lesion, which appears some 2–4 weeks later, is known as the chancre, and is found most often in the genitalia, lips and mouth. It consists of a painless, indurated papule that breaks down to form a typically flat hard ulcer (Figure 11.17) and heals completely even without treatment. The associated regional lymphadenopathy is also painless.

The disease becomes generalized (*secondary syphilis*) within 2–3 months of infection. A widespread skin eruption (papular, vesicular or bullous) develops predominantly on the face, palms and soles. In addition, other lesions such as condylomata lata, mucous patches and serpiginous ulcers occur, usually at mucocutaneous junctions. Constitutional symptoms include low-grade fever, sore throat, headaches, joint and muscle pain, generalized lymphadenopathy, iridocyclitis and anaemia. The disease remains highly infective during this stage. All secondary lesions heal spontaneously.

Tertiary syphilis is characterized by destructive lesions of a localized or diffuse nature that probably result from hypersensitivity to the spirochaetal antigens. The classic localized lesion is the gumma, which consists of an area of coagulative necrosis surrounded by a zone of lymphocytes, plasma cells and macrophages. Adjacent arteries exhibit marked endarteritis

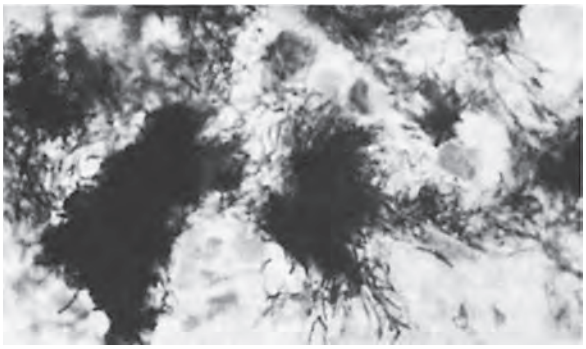


Figure 11.16 Microscopy of the sulphur granules from a patient with cervicofacial actinomycosis.

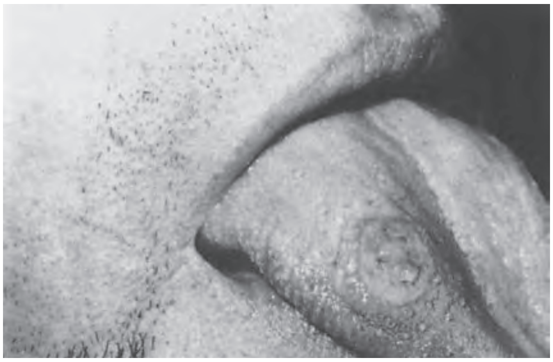


Figure 11.17 Syphilitic ulcer of the tongue: primary lesion.

obliterans. The most common sites of gumma include the testes, liver and bones (nose, palate, skull, clavicle, ulna and tibia). The bony lesions account for the deformities of tertiary syphilis, e.g. saddle nose (Figure 11.18). The diffuse tertiary lesions of syphilis include syphilitic aortitis and vasculitis, cerebral syphilis (meningovascular and parenchymatous) and diffuse syphilitic osteitis. The vascular lesions lead to weakening of the media with aneurysm formation and, in the case of the ascending aorta, with aortic regurgitation.

Meningovascular syphilis is characterized by focal meningitis, vascular episodes due to endarteritis and isolated cranial nerve palsies. Parenchymatous neurosyphilis comprises tabes dorsalis and general paralysis of the insane. In the former, there is degenerative demyelination and gliosis affecting the posterior columns of the spinal cord and posterior spinal nerve roots, resulting in the characteristic high-stepping gait. General paralysis of the insane is a chronic syphilitic meningoencephalitis. The disease affects the frontal lobes most severely.

Diffuse syphilitic inflammation of bones in tertiary syphilis is exemplified by the sabre tibia (Figure 11.19), in which the apparent bowing is due to the deposition of new periosteal bone, and the worm-eaten appearance of the skull that results from the combined effects of destruction and new bone formation.

Diagnostic tests

Treponema pallidum cannot be cultured. The organism can be identified from the exudate of primary and secondary lesions by dark-ground illumination after fluorescent antibody staining. Serological tests are based on either the Wasserman or the treponemal antibodies. The former include the Kahn and Venereal Disease Research Laboratories flocculation tests, and the latter the Reiter protein complement-fixation test, the *T. pallidum* immobilization test and the fluorescent treponemal antibody test. The serological tests using the treponemal antibody are more specific (fewer false-positive reactions).

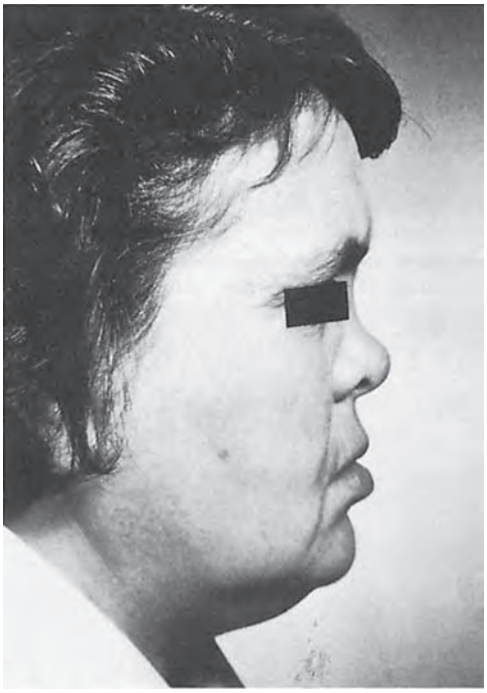


Figure 11.18 Saddle nose deformity in tertiary syphilis.

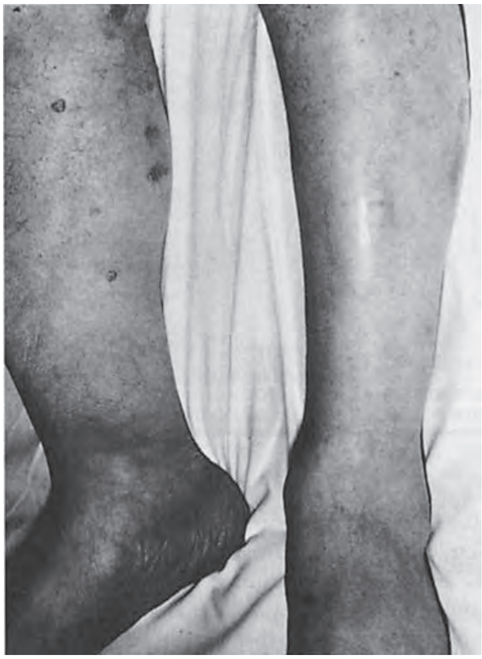


Figure 11.19 Diffuse syphilitic osteitis resulting in a sabre tibia.

Treatment

Both primary and secondary syphilis respond readily to adequate treatment, which is usually by intramuscular penicillin.

Gonorrhoea

This is caused by *N. gonorrhoea* and, in the vast majority of cases, transmission is by sexual intercourse.

Clinical features

In the female, the disease causes an acute purulent inflammation of the vulva, cervix, uterus and adnexa. Presentation with acute pelvic peritonitis and tubo-ovarian abscess is common. Secondary involvement of the rectum (proctitis) is found in 50% of females. Rarely, the rectum is the primary site of infection. In adult life the vagina is relatively resistant to gonococcal infection. This is not so in prepubertal girls, probably because of the immature and non-keratinized vaginal epithelium. Thus, gonococcal vaginitis and vulvovaginitis may occur in this age group. However, most cases of vulvovaginitis in prepubertal girls are not venereal in origin but result from endogenous infection with *Neisseria* spp. (*Neisseria sicca*, *Neisseria flava*, etc.) from the upper respiratory tract, the organisms being introduced into the vagina by dirty hands, towels, clothing, etc. Vulvovaginitis in both children and adults can be caused by *C. albicans* and *T. vaginalis.*

In the male, infection with *N. gonorrhoea* results in inflammation of the urethra but often also involves the epididymis, seminal vesicles, bladder and prostate. The untreated urethral inflammation may lead to stricture formation of the bulbar or spongy urethra.

In both sexes the disease may result in infertility owing to stricture of the fallopian tubes in the female and vas deferens in the male. One of the disastrous consequences of gonococcal infection in the female used to be infection of the eyeballs of the neonate during delivery (gonococcal ophthalmia neonatorum), which led to blindness in infancy. This is rarely encountered nowadays, with the advent of effective chemotherapy and routine instillation of silver nitrate or other antiseptic into the eyes of the newborn.

Treatment

Gonococcal infections are usually sensitive to penicillin. Tetracyclines are used for resistant infections, which account for 25% of cases.

Lymphogranuloma venereum

Lymphogranuloma venereum (LGV), also known as tropical bubo, is caused by *Chlamydia trachomatis.* Chlamydiae are obligate intracellular parasites and for this reason used to be considered as viruses. However, they have all the characteristics of bacteria, including a complex cell wall, but lack the metabolic enzymes necessary for an independent existence. *Chlamydia trachomatis* has three serotypes that cause LGV and others that are responsible for oculogenital infections.

LGV is common in tropical countries and is contracted by sexual intercourse, the reservoir of infection being the cervix in the asymptomatic female and the rectal mucosa in the asymptomatic homosexual male. It produces a papular, ulcerative or bullous lesion in the genital region that is not often painful and which heals spontaneously. This is followed within 1–6 weeks by gross lymphadenopathy (buboes). The enlarged lymph nodes in the ilioinguinal region suppurate and subsequently ulcerate, discharging seropurulent material. The disease may involve the pelvic organs and rectum in the female. It often becomes chronic with extensive scarring leading to elephantiasis and fibrous strictures of the rectum, vagina and urethra. Sulphonamide therapy (5 g daily for 7 days) gives good results in early cases. Tetracyclines are indicated for resistant cases. Surgical treatment should only be undertaken after an adequate course of chemotherapy. Abscesses should be aspirated and incision avoided.

Granuloma inguinale

This is found in certain tropical countries and the southern USA. It is an infection with Donovan bodies (donovania granulomatosis). Although infection is generally acquired by sexual intercourse, extragenital inoculation may occur. The primary lesion is a papule that ulcerates subsequently. In the male, it is usually found on the penis but other sites are well documented in both sexes. However, the majority of primary lesions occur in the genital, perineal, perianal or pubic regions. A 5 day course of streptomycin (4 g daily) is curative in most instances. Chlortetracycline, tetracycline and chloramphenicol are also effective.

Chancroid (soft chancre)

This is caused by *Haemophilus ducreyi* and is transmitted by sexual contact. The infection is more common and more severe in males. The disease starts as a soft macule, usually in the foreskin, 3–10 days after exposure. The lesion subsequently becomes necrotic and produces a ragged ulcer that may result in substantial penile destruction. These ulcers may be multiple and vary considerably in size. In the female, ulcerative lesions are found in the vulva and vagina. The inguinal lymph nodes become enlarged and painful and may suppurate. Treatment is by sulphonamides and tetracycline in the first instance. Resistant cases respond to cephalothin.

Condylomata acuminata

The term 'condylomata acuminata' (genital warts) is used to differentiate these pointed warts from the flat condylomata lata of syphilis. Genital warts result from infection with the human papillomavirus (HPV) and occur in the genital, perineal and perianal regions (Figure 11.20), and may be followed by the appearance of warts elsewhere in the body. Treatment is by the application of 10% podophyllin or excision. See HPV later in this chapter.

Viral infections of surgical importance

The important infections in surgical practice are the Herpetoviridae, HPV, viral hepatitis and AIDS.

Herpetoviridae infections

This family comprises a large number of viruses but only four are pathogenic to humans. These are the herpes simplex virus (HSV), varicella zoster virus (VZV), cytomegalovirus (CMV) and Epstein–Barr virus (EBV). They are important because of their common and ubiquitous occurrence, and because of their association in some instances with the development of certain neoplastic conditions.

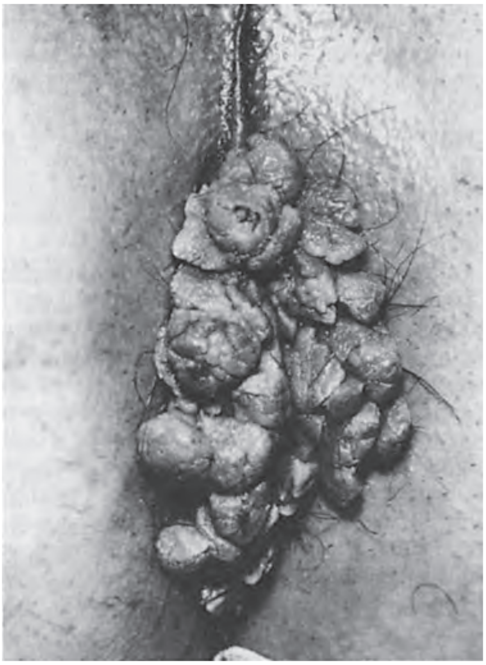


Figure 11.20 Condylomata acuminata of the perianal region.

Moreover, they can cause serious and at times fatal infections in debilitated and immunocompromised patients.

These viral infections cause an initial, often mild and inconsequential, primary infection, following which the virus remains dormant in a non-infectious state (latent) at certain sites, e.g. sensory ganglia in HSV and probably in VZV. From time to time, reactivation of the virus with occurrence of clinical manifestations, such as cold sores or shingles, may follow a febrile illness, operation, menstruation, radiotherapy, etc. The latency site for the EBV genome is thought to be the B-lymphocytes, whereas the CMV genome appears to persist in the renal tubules, leucocytes (particularly polymorphs), parotid gland and cervix.

Herpes simplex infections

There are two recognized strains: HSV-1, which is responsible for the majority of non-genital infections, and HSV-2, which accounts for most of the ocular and genital infections. HSV infections are widespread and prevalence rates range from 50% to 100% depending on socioeconomic status. The infection is acquired in the first instance by close personal contact, and genital herpes is now one of the well-recognized venereal diseases in the West, with a rising incidence such that in the UK genital herpes is twice as common as syphilis.

Labial herpes

This is the most common type of infection caused by HSV-1. The portal of entry is the mouth, and vesicular lesions occur most commonly on the lips, as cold sores, but may be severe, especially in the neonate, debilitated and immunocompromised, in whom they may cause extensive gingivostomatitis with intraoral/pharyngeal ulceration, fever and cervical lymphadenopathy. Rarely, the infection may disseminate to the brain, liver and adrenals, and is then usually fatal. Labial herpes is the most common recurrent type. Precipitating factors include fever, exposure to sunlight and menstruation. Each crop of labial vesicles is preceded by a burning sensation in the skin of the affected site.

Cutaneous herpes

Although HSV does not penetrate intact healthy skin, infections may occur in patients with skin disorders and in the presence of burns and lacerations leading to an extensive vesicular eruption (eczema herpeticum). In burn patients, the infection may become widely disseminated and lead to severe and often fatal pneumonia.

Ocular herpes

Most of these infections are due to HSV-1 except for those acquired in the neonatal period from an infected mother. The eye lesions include follicular conjunctivitis that may progress to corneal involvement with the formation of dendritic ulcers and corneal opacities. Corticosteroids enhance the eye damage caused by the virus and should be avoided in these infections.

Central nervous system infections

HSV infections may cause encephalitis, radiculitis, myelitis and meningitis. The encephalitis is usually the result of reactivation of latent infection.

Genital herpes

Although the majority of genital herpes infections are due to HSV-2, some are caused by HSV-1. The disease, which is usually sexually transmitted, has an incubation period of 2–20 days with an average of 6 days. The primary infection is followed by recurrent attacks that are usually less severe with intervening periods of latency during which the patient feels well and has no clinical manifestations. The risk of transmission of the disease is highest during an acute attack but asymptomatic individuals may pass on the disease to their sexual partners. There is strong evidence linking herpes simplex infection of the cervix with the development of cervical cancer and a fourfold increased risk is encountered in females with genital herpes. An association between cancer of the vulva and genital herpes has also been suggested, although the evidence linking the two conditions is inconclusive at present.

A primary attack is heralded by systemic symptoms due to the viraemia: malaise, fever and myalgia. The lesions are found on the penis and perianal region in the male, and labia, clitoris, vagina, cervix, perineum and perianal region in the female. They consist of painful vesicles that ulcerate, then crust and heal spontaneously. Neuralgia is often present. The severe pain during an attack may precipitate acute retention of urine.

Treatment of herpes simplex virus infections

As the virus is metabolically inactive during a latency period, treatment is futile at this stage. Effective therapy with the antiviral agent aciclovir is possible provided it is started early during a first or recurrent attack. There is a preferential uptake of the drug by the infected (virally colonized) cells, where it is activated by a virus-specific thymidine kinase forming aciclovir triphosphate, which inhibits the viral DNA polymerase.

Varicella zoster virus infections

The painful clinical condition known as zoster (formerly herpes zoster) is always the result of a reactivation of the same virus that causes chickenpox. Although in some instances no apparent cause for this activation is clinically obvious, in others a state of depressed cell-mediated immunity is present due to trauma, malignancy and immunosuppressive drugs, particularly in transplant patients. In severely immunocompromised patients, the infection may be systemic, with the development of pneumonia and involvement of other organs such as the liver, CNS, adrenals and pancreas. The systemic disease carries a very high mortality.

In the more usual condition, an exanthematous rash develops over a dermatome supplied by a specific dorsal nerve root or extramedullary nerve ganglion. The dermatomes supplied by the third dorsal to the second lumbar segment of the spinal cord are the ones most commonly affected, followed by that supplied by the fifth cranial nerve. Pain and paraesthesiae often predate the development of the vesicular rash. The most distressing feature of the disease is the development of neuralgia (St Anthony's fire), which may persist for several weeks and require specialist treatment.

Cytomegalovirus infections

The pattern of infection with CMV varies with the socioeconomic state of the country. In poor, developing countries with overcrowding and poor sanitation, the disease is acquired early in life and, by the age of 5 years, the vast majority of children become seropositive for the virus. In the West neonatal infection is rare and most primary infections are acquired in adolescence, mainly from kissing and sexual intercourse, such that the prevalence of seropositive individuals rises to 60–70% by the age of 60 years. The disease can be transmitted by the transfusion of blood products, particularly fresh blood and pooled platelet donations, and organ and bone marrow transplantation. It is thus a real hazard in transplant patients and in patients with leukaemia, since the transmission of the virus may be followed by serious infection owing to the immunocompromised state, leading to hepatitis, haemolytic anaemia, leucopenia and thrombocytopenia. The average reported mortality of this generalized CMV disease is 2% but can be as high as 15–20%, especially in bone marrow transplant recipients.

The infection may be transmitted to the fetus across the placenta. This may result in intrauterine death and spontaneous abortion. The vast majority of congenitally infected babies appear normal at birth, but 10–30% of them will suffer brain damage (microcephaly) and mental retardation. A few babies show the classic cytomegalic inclusion disease, the features of which are similar to those of the generalized disease encountered in immunosuppressed adults.

Infection in a normal adolescent may be asymptomatic or the individual may develop fever, sore throat, lymphadenopathy and hepatosplenomegaly. The clinical picture is very similar to that of glandular fever but the Paul–Bunnell test is negative. The blood may contain the characteristic intranuclear inclusions (owl eye) within atypical mononuclear cells. Anti-CMV antibodies can be demonstrated following the primary infection.

In susceptible immunocompromised patients, the risk of CMV infection may be reduced by using blood and organs from CMV-negative donors, transfusion of leucocyte-free blood and administration of high-titre anti-CMV immunoglobulins to the recipients.

Epstein–Barr virus infections

This gammaherpes virus was the first virus to be identified in human neoplastic cells. In excess of 90% of the world population becomes infected with EBV before adolescence. Aside from causing glandular fever (infectious mononucleosis), infection with EBV leads to the development of a number of tumours, e.g. lymphoproliferative disorders (endemic Burkitt's lymphoma, opportunistic lymphoma, nasal natural killer cell lymphoma), gastric carcinoma (10–15% of cases worldwide), nasopharyngeal carcinoma and spindle cell tumours (in immunocompromised hosts). Individuals who become infected with EBV become life-long carriers. The virus homes on two target cell types – B-lymphocytes (latent virus) and the epithelial cells of the oropharynx (replicating virus) – which are thus the main site of intermittent production of infectious virus, the B-lymphocytes becoming infected by circulating within the oropharyngeal mucosa. The infected B-lymphocytes then carry the virus elsewhere, with infection of other epithelial sites.

Infectious mononucleosis is a disease of the West with a peak incidence in adolescence and young adult life. It is transmitted in the saliva by kissing and is often referred to as the 'kissing disease'. The incubation period varies from 4 to 7 weeks and is followed by the development of malaise, fever, asthenia, sore throat, lymphadenopathy and splenomegaly. The blood picture shows a lymphocytosis with more than 10% atypical monocytes. The Paul–Bunnell test, which detects the presence of heterophile antibodies to sheep red cells, is positive. The liver function tests are often deranged during the first week of the illness, which usually lasts for 3–4 weeks. However, prolonged asthenia and debility for several months after the acute illness are quite common.

Human papillomavirus

HPVs are DNA viruses that infect the skin and squamous epithelium of the mouth and throat, the cervix and the anus. There are more than 150 different types of HPVs designated by a specific number. HPV infection is very common and most people have HPV infection at some time during their lives. In the majority the infection is asymptomatic and is cleared by the body defence system. HPV infections are usually acquired at an early age before the development of specific immunity. Some types of HPVs, if persistent, cause changes in the cervix, lips, buccal, oropharyngeal and anal mucosa and collectively are known as high-risk HPVs because they initiate dysplastic changes that may progress to *in situ* and invasive cancers. It has been estimated that HPV infection accounts for approximately 5% of all cancers worldwide.

Other types of HPVs cause warts, verrucas and perianal warts (condylomata); these are often referred to as the 'wart or genital wart viruses' and include types 6 and 11. Genital warts may appear within several weeks after sexual contact with a person who is infected with HPV, or they may take months or years to appear. They produce warts and verrucas most commonly on the hands and feet, and in the genital and perianal areas, but they can occur elsewhere in the body. These HPVs do not usually cause precancerous lesions, and for this reason are referred to as low-risk HPVs. Infection of low-risk HPVs is by skin contact with an affected person. Common warts are caused by HPV types 1 and 2, whereas genital warts are commonly caused by types 6 and 11, although several other types may be responsible. HPV infections of the mouth and pharynx are usually through kissing and genital HPV is transmitted by unprotected sex.

Human papillomavirus and cervical cancer

The high-risk HPVs that increase the risk of cervical cancer are types 16, 18, 31, 33 and 45. Almost all women who develop cervical cancer have at least one of these types of HPV in the cells of their cervix. Types 16 and 18 cause approximately 70% of cancers of the cervix; the other types account for the remaining 30% of cervical cancers. Other factors, in addition to persistent infection with these high-risk HPVs, are involved in the development of cervical cancer in the individual patient and these include the immune status, smoking and multiparity. Individuals with low immunity from any cause (including drug induced) have an increased risk of cervical cancer. HPV DNA testing is now available in most countries and, since April 2011, it has become part of the UK National Health Service Cervical Screening Programme. HPV DNA tests look for viral DNA before any cell abnormalities become visible. The test is offered to women with borderline or low-grade cervical changes detected in the first round of cervical screening. Women who test positive for high-risk types of HPV are referred for a colposcopy. Those who are negative for HPV will only be offered monitoring to establish whether the cervical changes will resolve within a few months. There is no treatment for eradicating the HPV virus. The risk of cervical HPV infection is reduced but not abolished by protected sex with use of a condom. Vaccines are now available to prevent infection with types of HPV that can lead to cervical cancer. These include the Gardasil and Cervarix cervical cancer vaccines, which have been licensed in the UK since 2007. These vaccines should prevent this type of cancer in the future. Currently, all girls aged 12 or 13 in the UK are offered the cervical cancer vaccine subject to approval by their parents.

Human papillomavirus and other cancers

Some types of HPV increase the risk of developing other cancers: anal cancer, cancer of the penis, vaginal and vulval cancer and oropharyngeal cancers (middle part of the throat, including the soft palate, the base of the tongue, and the tonsils). HPV type 16 causes approximately 85% of anal cancers. Some oropharyngeal cancers are caused by HPV infection, almost exclusively by HPV type 16.

Treatment of lesions caused by human papillomavirus infections

Methods commonly used to treat cervical lesions include cryosurgery, electrosurgical excision procedure and conization (removal of cone-shaped areas of tissue containing the dysplastic epithelium from the cervix and cervical canal). Warts are treated by excision, fulguration or cryotherapy.

Viral hepatitis

There are many viruses that can infect the human liver and cause hepatitis of varying severity (Table 11.2). Some are essentially hepatotropic, i.e. the liver is the primary site of infection.

At least six distinct viruses fall into this category: A, B, C, D, E and G. In addition, others are still undefined and are referred to as non-A and non-B (NANB).

Viral hepatitis is the most common liver disease today and constitutes a worldwide problem. Following the acute illness that may progress to fulminant liver failure, persistence of the virus leads to chronic liver disease, the severity of which varies from healthy asymptomatic carrier to ongoing chronic active disease with progression to cirrhosis and, in the case of hepatitis B and C disease, the development of hepatocellular carcinoma.

Hepatitis A

Hepatitis A virus (HAV) is also known as enterovirus type 72 and the disease was formerly known as infectious hepatitis. The virus is shed in large numbers in the faeces for a few weeks before the overt clinical illness, which varies in severity. The immunoglobulin (Ig) M antibody to the virus (anti-HAV) is present in high titres in the serum early on during the acute disease and is detected for 6 months after an acute infection. The IgG antibody indicates past infection and is present in 40–50% of the urban population in the UK. Hepatitis A rarely gives rise to chronic liver disease and carries a low mortality.

There are two vaccines (Havrix and VAQTA) which are used in the prophylaxis against hepatitis A. Both contain no live virus and are very safe. In addition, combined vaccine (Twinrix) is available against both hepatitis A and B and is used in adults (individuals over age 18). Protection from HAV starts about 2–4 weeks after the first injection. A second dose is necessary to ensure long-term immunity. Vaccination against HAV is recommended in the following:

- All children older than 1 year are recommended to get the vaccine if they live in communities where the prevalence of HAV infection is high or where there are periodic outbreaks of hepatitis A.

Table 11.2 Viruses causing acute hepatitis and chronic liver disease

Virus	Transmission route
Hepatitis A	Enteral transmission, rarely causes liver cell necrosis
Hepatitis B (HBV)	Transmitted by blood products, needles, tattooing, sexual activity, mothers to babies, aerosol (dental treatment); can progress to liver cell necrosis and chronic liver disease
Hepatitis C	Transmitted sexually and by blood products, can progress to liver cell necrosis; high incidence of progression to chronic liver disease
Hepatitis delta	Incomplete hepatotropic virus; capable of infection only when activated by HBV. Acquired either as a coinfection with HBV or as superinfection in HBsAg carriers
Hepatitis E	Enteral transmission. Mild self-limiting disease. Does not progress to chronic liver disease
Hepatitis G	Transmitted by blood products. Uncertain clinical significance
Epstein–Barr virus	Agent of infectious mononucleosis. Hepatitis rare and usually mild
Cytomegalovirus	Immunosuppressed patients and infants
Yellow fever virus	Can cause liver cell necrosis when mortality is high
Ebola and Marburg virus	Causes African haemorrhagic fever, consisting of papular rash, DIC, pancreatitis and hepatitis. Spread by needles and person to person. High mortality. No specific therapy available
Others	

DIC, disseminated intravascular coagulation; HBsAg, hepatitis B surface antigen.

- Individuals who are likely to be exposed to HAV at work: staff in research virology laboratories, healthcare workers, food service workers, day care personnel, and sewage and waste-water workers.
- Individuals travelling to developing countries known to have a high prevalence of HAV; vaccination advised at least 4 weeks before travel.
- Active practising male homosexuals.
- Drug addicts.
- Individuals with impaired immunity and/or chronic liver disease.
- Haemophilic patients needing clotting factors therapy.

If an uninfected person is exposed to hepatitis, treatment with immune serum globulin is advised and if administered within 2 weeks after exposure may prevent infection.

Hepatitis B

Hepatitis B virus (HBV) infection is caused by a double-stranded DNA virus that replicates by reverse transcription. Hepatitis B is endemic in the entire human population and hyperendemic in many parts of the world. The extent of the infection is indicated by the estimated 300 million people who have persistent infection with the virus. The transmission of the acute disease is by the parenteral route and was formerly common as a complication of blood transfusion; hence, the old terminology of 'serum or post-transfusion hepatitis'. With the testing of donors for the HBV, and the heat-inactivation methods for the preparation of the protein fractions, hepatitis B after transfusion of blood and blood products has been virtually eliminated. The HBV has great infectivity and is transmitted by needle sharing in drug misusers, in whom it is prevalent. It has caused disease in surgeons and nurses as a result of accidental injuries by 'sharps'.

The HBV has been fully characterized. The whole virus (virion) is known as the Dane particle and consists of a core covered by surface antigen [hepatitis B surface antigen (HBsAg)]. The latter is formed in excess as separate tubules and spheres. The centre of the virus contains a core antigen (HBcAg) that is never found in the circulating blood, the 'e' antigen (HBeAg), double-stranded DNA and a DNA polymerase. Antibodies are formed to the various antigens and can be detected in the peripheral blood at some time in all patients during the acute infection and in some carriers. The HBeAg is encountered early on during the course of the acute hepatitis, and is usually associated with high titres of HBsAg and the presence of Dane particles (complete HBV). It is the best marker of probable infectivity, and is usually replaced by the anti-HBeAg antibody soon after the jaundice develops. Anti-HBsAg denotes recovery from acute infection and immunity. Anti-HBcAg is detected in the peripheral blood early on during the acute infection. It is not protective and is found in all carriers. Anti-HBeAg is present in convalescent patients and in the majority of carriers. This antibody is associated with a low probability of infectivity. The interpretation of the various serum markers of hepatitis B is shown in Table 11.3.

Hepatitis B may be followed by chronic liver disease, which in some types progresses to cirrhosis and end-stage liver disease requiring liver transplantation. It is also responsible for the development of hepatocellular carcinoma. The viral genome (DNA) of HBV becomes incorporated in the host DNA and this integration has been shown to be a necessary step in the development of primary hepatocellular carcinoma. The important role of HBV infection in the development of hepatocellular carcinoma is now well established in hyperendemic areas.

Chronic hepatitis B

Chronic infection with HBV occurs in 5–10% of adult patients who are unable to eliminate the HBsAg within 3 months from the acute infection. The previous classification of chronic hepatitis B into the healthy HBsAg carrier state and patients with chronic B hepatitis (chronic persistent and chronic active) has been superseded by a more valid classification that has therapeutic and prognostic implications. The new classification is based on (1) the presence or absence of HBV replication (assessed by measuring HBV DNA or HBeAg in the serum, or HBcAg in the liver) and (2) the presence or absence of liver cell inflammation by assay of the serum aminotransferase activities, ASAT (ALT) or SGOT. Replication of HBV is usually associated with raised ASAT or SGOT. Three categories are thus recognized:

1. chronic HBeAg-positive hepatitis (CHBe^{+})
2. chronic HBeAg-negative hepatitis with normal aminotransferase activity (CHBe^{-}/ASAT^{-})
3. chronic HBeAg-negative hepatitis with raised aminotransferase activity (CHBe^{-}/ASAT^{+}).

Chronic HBeAg-positive hepatitis

Antigen-positive disease with evidence of viral replication constitutes 15% of patients. The disease is usually associated with continued liver cell inflammation (raised transaminases) and is symptomatic. On follow-up seroconversion from HBe^{-} to HBe^{-} occurs in about 50% of patients during the ensuing 5 years, with improvement in symptoms and normalization of the liver function tests. However, subsequent reactivation (with seroconversion to HBe^{-}) occurs in 5% of patients. The overall 5 year mortality is 10%.

Chronic HBeAg-negative hepatitis with normal aminotransferase activity

The 'normal carrier' state accounts for the majority of cases (60%). These patients have no viral replication or symptoms and their liver function tests are normal. The outcome of this group of patients is good and the mortality is low but they are subject to superinfection with hepatitis delta virus (HDV) (see below). The reactivation with seroconversion to HBe^{-} is low in these patients.

Table 11.3 Guide to the interpretation of serum markers of hepatitis B

HBsAg	HBeAg	Anti-HBe	Anti-HBc	Anti-HBs	Interpretation
+	+	–	–	–	Incubation period or early acute hepatitis B
+	+	–	+	–	Acute hepatitis B or persistent carrier
+	–	+	+	–	Late acute hepatitis B of persistent carrier
+	–	–	+	–	Late acute hepatitis B or persistent carrier
–	–	+	+	+	Convalescent acute hepatitis B
–	–	–	+	+	Past infection

HBeAg; HBsAg, hepatitis B surface antigen.

Chronic HBeAg-negative hepatitis with raised aminotransferase activity

This is encountered in 25% of patients and forms a heterogeneous group with either ongoing HBV replication (liver HBcAg$^-$ despite being seronegative for HBeAg) or other hepatic disease, e.g. concomitant HDV infection, autoimmune reactivity, alcohol misuse, drug-related hepatitis, hepatocellular carcinoma or schistosomiasis. The outcome in the individual patient depends on the exact aetiology. Symptomatic decompensated cirrhosis is highest in this category and develops in 20–30% of patients.

Drug treatment of chronic hepatitis B

Tenofovir disoproxil is a nucleotide analogue. It works by blocking the enzyme reverse transcriptase, which is responsible for HBV replication. Tenofovir disoproxil is used for the treatment of chronic hepatitis B in adults with compensated liver disease, with evidence of active viral replication, persistently elevated serum alanine aminotransferase (ALT) levels and histological evidence of active inflammation and/or fibrosis. Adverse events associated with the use of nucleotide analogues include lactic acidosis and progression of hepatomegaly. Other side effects reported for tenofovir disoproxil include headache, fatigue and gastrointestinal disorders.

Vaccination against hepatitis B

There is now an official recommendation that health service workers, particularly all doctors and nurses, should undergo active immunization against hepatitis B. Within the UK all students seeking admission to medical schools must also undergo this vaccination. The licensed vaccines that are available are prepared either by means of genetic engineering (recombinant hepatitis B vaccine) or from fully purified formalin-inactivated B surface antigen (HBsAg) obtained from the plasma of known carriers. Both types of vaccine are safe and effective. The vaccination programme consists of three doses administered at zero time, 2 months and a booster at 6 months. This regimen is effective in preventing the development of hepatitis B if immunization is started within a few weeks of exposure, that is, after an interval which is less than the incubation period of the disease. It is the recommended treatment in non-immunized medical and nursing staff who sustain accidental injury by 'sharps' contaminated with blood from an infected patient. Some also advise the administration of hepatitis B–specific immunoglobulin for additional (passive) immunization. The previous concern that some of these commercial IgG preparations were unsafe because of contamination by HIV has been abolished by HIV testing of donors and the introduction of rigorous techniques that ensure viral inactivation during the processing of the donated plasma.

Hepatitis C

The hepatitis C virus (HCV) is a small (40–60 nm) bilipid enveloped, single-stranded RNA virus of the family Flaviviridae and genus *Hepacivirus* (Figure 11.21). There are various major genotypes and more than 50 subtypes of HCV depending on geographical distribution of the virus. HCV is spread primarily by contact with blood and blood products. Since the introduction of routine blood donor screening for HCV, transfusion-related hepatitis C has virtually disappeared. At present, injection drug use is the most common risk factor, although many patients acquire hepatitis C without any known exposure to blood or to drug use. Aside from drug users and patients who had blood transfusions or blood products before donor testing (1992), e.g. haemophiliac and transplant patients, the other at-risk group consists of blood-borne infection of babies of HCV-infected mothers and healthcare workers who sustain needle-stick injuries. There is also some evidence that hepatitis C is common in promiscuous individuals with multiple sexual partners and cocaine addicts who use shared equipment for intranasal administration. However, in 10% of patients who contract acute hepatitis C, the cause is not known as there are no obvious known risk factors. These are referred to as sporadic or community-acquired infections and they make up 30% of patients with chronic hepatitis C.

HCV is a major cause of chronic liver disease and hepatocellular carcinoma worldwide. Following the acute infection, 85% of patients develop chronic infection with persistent viraemia. Chronic hepatitis may run a mild course with fluctuating levels of transaminase levels over several years, but progression to cirrhosis is encountered in 30% of cases. The variable course is thought to be the result of host factors, e.g. route and size of the infecting inoculum, patient's sex, age at time of infection and alcohol consumption. Thus, severe chronic disease is rare in females, whereas chronic alcohol consumption and increasing age at the time of infection are strong risk factors for disease progression. Cirrhosis due to HCV is now the most common indication for liver transplantation in the West. Recurrence of HCV in the grafted organ is very common and 50% of transplanted patients develop chronic hepatitis. Early intervention with interferon and ribavirin may prevent or delay the progression of HCV-related graft disease. As with HBV, there is a strong link between persistent HCV infection and the development of hepatocellular carcinoma. Thus, anti-HCV antibodies are found in over 70% of non-hepatitis B patients who develop this tumour.

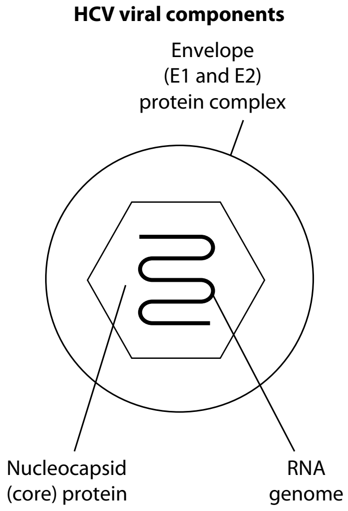

HCV viral components
Envelope (E1 and E2) protein complex
Nucleocapsid (core) protein
RNA genome

Figure 11.21 Hepatitis C virus (HCV).

Clinical features

Many patients have asymptomatic disease or suffer from mild, non-specific symptoms, e.g. fatigue, mild right upper quadrant discomfort, nausea, poor appetite and myalgia. Likewise, physical examination is normal except for mild hepatic enlargement/tenderness. Those with established liver disease may exhibit the symptoms and signs of cirrhosis: hepatomegaly, jaundice, splenomegaly, fluid retention, etc. A small percentage of patients with chronic hepatitis C develop cryoglobulinaemia, vasculitis and glomerulonephritis.

Acute hepatitis C is diagnosed by (1) marked (more than 10 times upper limit) elevation of serum aminotransferases; (2) positive anti-HCV is present in serum; and (3) its confirmation by the presence of HCV RNA in serum. *Chronic hepatitis C* is diagnosed when anti-HCV is present and serum aminotransferase levels remain elevated for more than 6 months. Testing for HCV RNA (by PCR) confirms the diagnosis as almost all patients with chronic infection will have the viral genome detectable in serum.

Serological testing for anti-HCV antibody by enzyme immunoassay (EIA-3) is the initial test but in view of false-positive results that may occur with EIA-3 testing, confirmatory testing with assay for HCV RNA by PCR or transcription-mediated amplification is necessary for confirmation of diagnosis. A positive HCV RNA in serum indicates active infection. Testing for HCV RNA is always necessary in immune-suppressed/-compromised patients. The anti-HCV serological test (EIA-3) may be negative in the early stages of acute HCV infection but becomes positive in almost all patients within 2–8 weeks of onset of acute illness. Western blot assays (immunoblot assays) are also used to confirm anti-HCV. Immunoblot tests are used routinely in blood transfusion departments for further testing of blood donor samples positive on screening with EIA-3. Quantitative tests for HCV RNA in serum are used as an indirect assessment of viral load and are based on quantitative PCR and a branched DNA testing. Most patients with chronic hepatitis C have levels of HCV RNA (viral load) between 100 000 and 10 000 000 copies/mL. Although viral loads do invariably correlate with the disease severity or poor prognosis, they are useful in the identification of patients who are likely to respond to antiviral therapy. Liver biopsy is not necessary for diagnosis but is helpful for grading the severity of disease based on extent of liver parenchymal damage: periportal inflammation/fibrosis, piecemeal necrosis and established cirrhosis. Specific immunohistochemical stains for HCV have not been developed. Liver biopsy is also helpful in ruling out other causes of chronic liver disease.

Genotyping and serotyping of hepatitis C virus

Aside from varying with the geographical location and thus being useful in analysis of epidemic outbreaks, these tests are important as they impact on management, particularly on medical treatment. Thus, patients with genotypes 2 and 3 are more likely to respond to interferon-based therapy than patients with genotype 1 and require only a 6 month course of combination treatment using interferon and reduced daily dose (800 mg) of ribavirin, whereas patients with genotype 1 require a longer course (12 months) and a full dose of ribavirin (1000–1200 mg daily).

Treatment

Currently, the optimal regimen is a 24 or 48 week course of the combination therapy consisting of pegylated α-interferon and ribavirin. Pegylated α-interferon therapy, which has replaced other recombinant forms of α-interferons, is complexed with polyethylene glycol, which prolongs its half-life, thereby enabling weekly injection. Moreover, pegylated interferons are more active than standard interferons in inhibiting HCV and yield higher sustained response rates.

Ribavirin is an oral antiviral agent that has activity against a broad range of viruses. By itself, ribavirin has little effect on HCV, but, as combined therapy with interferon, it increases the response and viral clearance rates. Ribavirin causes certain side effects, including anaemia, fatigue and irritability, itching, skin rash, sinusitis and cough. It also causes a dose-related haemolysis of red cells. For these reasons, ribavirin should not be used in patients with pre-existing anaemia or in patients with significant coronary or cerebral vascular disease. If such patients require therapy for hepatitis C, they should receive α-interferon monotherapy. Ribavirin has been shown to cause birth defects in animal studies and is thus contraindicated in pregnancy and in women/couples who are unwilling/unable to practise contraception.

Combination therapy, when possible, is the standard treatment for hepatitis C, and interferon monotherapy is indicated only when ribavirin is contraindicated. The two forms of pegylated interferon used clinically are peg-interferon α-2a (Pegasys) and peg-interferon α-2b (Pegintron). Both are administered subcutaneously at weekly intervals but in different dosage: peg-interferon α-2a in a fixed dose of 180 μg/week and peg-interferon α-2b in a subcutaneous dose of 1.5 μg/kg. Ribavirin is administered orally in a dose of 200 mg b.d. up to a total administered dose of 1.0 g in patients weighing less than 75 kg, and 1.2 g when weight exceeds 75 kg. In certain situations, an 800 mg dose (400 mg twice daily) is recommended (see below).

Combination therapy leads to rapid improvements in serum ALT levels and disappearance of detectable HCV RNA in up to 70% of patients. Long-term improvement in hepatitis C occurs only if HCV RNA disappears during therapy and remains undetectable once therapy is stopped. Some patients relapse after treatment is stopped but the relapse rate is less with combination therapy.

For patients treated with combination therapy, the optimal duration of treatment depends on viral genotype. Patients with genotypes 2 and 3 have a high rate of response to combination treatment (70–80%), and a 24-week course of combination therapy is sufficient. In contrast, patients with genotype 1 have a lower response rate to combination therapy (40–45%), and a 48-week course is necessary. The optimal dose of ribavirin also varies with the genotype, 800 mg daily being adequate for patients with genotypes 2 or 3 but a full dose (1.0–1.2 g depending on body weight) being required to achieve control of the viral load.

Combination medical therapy is indicated in all patients with anti-HCV, HCV RNA, elevated serum aminotransferase levels, and evidence of chronic hepatitis on liver biopsy in the absence of contraindications. In essence, combination therapy for hepatitis C is indicated in all patients who have histological evidence of progressive disease. Patients with less severe histological disease are managed on an individual basis. Patients with cirrhosis are candidates for combination therapy only if they have compensated cirrhosis (Child A disease). Elderly patients (>60 years) have a higher incidence of adverse side effects but are still candidates for combination therapy on an individual basis. Patients with both HCV and HIV infection should be offered therapy for hepatitis C as hepatitis C tends to be more rapidly progressive in patients with HIV coinfection. However, treatment for hepatitis C is withheld in patients with advanced HIV disease as complications of therapy are common and response rates are low. In patients with clinically significant extrahepatic manifestations, such as cryoglobulinaemia and glomerulonephritis, therapy with α-interferon can result in remission.

Combination antiviral treatment for hepatitis C is recommended in:

- patients at increased risk for cirrhosis (persistent elevation of ALT, high levels of HCV RNA in the blood, evidence of early fibrosis, moderate to severe inflammation and piecemeal necrosis on liver biopsy)
- patients who are coinfected with HCV and HIV, because these patients have a more rapid progression of the disease
- persistent cryoglobulinaemia.

Combined antiviral therapy is inadvisable in:

- patients with non-progressive disease and near normal ALT
- patients who are unable to comply with the treatment schedule
- patients in whom treatment is unsafe (for example, allergy to the medications)
- transplant patients
- pregnant women and females unwilling to practise adequate contraception during treatment, or who have reversible serious untreated conditions such as unstable heart disease
- uncontrolled high blood pressure
- untreated major depression.
- patients with decompensated cirrhosis (Child C disease)
- clinically decompensated cirrhosis due to hepatitis C.

α-Interferon can induce adverse neuropsychiatric effects, including marked irritability, anxiety, personality changes, depression, and even suicide or acute psychosis. α-Interferon therapy can also induce formation of autoantibodies, especially in patients with an underlying susceptibility to autoimmunity. Exacerbation of a known autoimmune disease, e.g. rheumatoid arthritis or psoriasis, is common during interferon therapy. Contraindications to α-interferon therapy include:

- severe depression or other neuropsychiatric syndromes
- drug or alcohol abuse
- autoimmune disease that is not well controlled
- patient with bone marrow compromise
- females unable/unwilling to practise birth control.

Contraindications to ribavirin are:

- marked anaemia
- renal dysfunction
- coronary artery or cerebrovascular disease
- females/partners unable/unwilling to practise birth control.

All patients with HCV should be vaccinated against hepatitis B and hepatitis A. They also should be counselled on measures to prevent the spread of HCV and advised to abstain from alcohol use. Finally, all patients with HCV should be tested for HIV.

Hepatitis delta virus disease

HDV is an incomplete hepatotropic virus that is unable to replicate on its own. It is thus capable of infection only when activated by the presence of HBV, when it forms a virion particle consisting of an outer coat (from the HBs antigen of HBV) and an inner core of HDAg and RNA genome. Estimates indicate that more than 15 million people are infected worldwide. There are eight reported genotypes of HDV with varying pathogenicity. Areas with the highest prevalence include southern Italy, North Africa, the Middle East, the Amazon Basin and the American South Pacific islands. HDV infection is more common in adults than in children.

The disease may be acquired either as a coinfection together with HBV in previously normal individuals or as superinfection in carriers of HBsAg. Special subsets of HBsAg carriers are at high risk of superinfection. These include those individuals who are exposed to multiple blood and interpersonal contacts such as intravenous drug addicts, haemophiliacs, institutionalized patients, prisoners and homosexuals. Parenteral inoculation appears to be the most efficient mode of transmission. It is now known that blood that is negative for the HBsAg may transmit HDV at a frequency of 1 in 3000 transfusions.

Coinfection in previously normal individuals usually leads to a self-limiting hepatitis with a variable clinical presentation. Both IgM (during acute illness) and IgG (in the convalescent period and for 1–2 years after clearance of the HBsAg) anti-HDAg antibodies appear in the plasma in moderate to severe disease, which may progress to fulminant hepatic failure. The more usual mild form of coinfection causes a minimal serological response. Coinfection rarely progresses to fulminant liver failure (1% of patients) but complete clinical recovery and clearance of HBV and HDV are the most common outcomes.

Infection with HDV in a patient who is already positive for the HBsAg is known as superinfection and results in severe fulminant liver failure in 5% of patients and approximately 80–90% develop chronic HDV infection. These patients progress more rapidly to develop cirrhosis and may develop hepatocellular carcinoma. In superinfection, the HDV aggravates pre-existing type B hepatitis (chronic aggressive hepatitis) or induces new HDV disease in healthy HBsAg carriers. In the latter group, the hepatitis is often severe and may progress to fulminant hepatic failure.

Treatment consists primarily of support. Patients with evidence of decompensated liver disease or fulminant liver failure should be immediately transferred to a transplant centre.

Currently there is no effective treatment for HDV. However, a study of the efficacy of peg-interferon α-2a found that treatment with or without adefovir over 48 weeks resulted in sustained HDV RNA clearance in approximately 25% of patients.

Hepatitis E

Hepatitis E virus (HEV) consists of a non-enveloped single-stranded RNA, which causes enterally transmitted hepatitis. The epidemiological features are similar to those of hepatitis A but the E virus is serologically different and it can infect individuals who have recovered from hepatitis A. The disease is spread by contamination of drinking water with sewage. Risk factors for HEV infection include poor sanitation and HEV shedding in faeces. Person-to-person transmission is uncommon. There is no evidence for sexual transmission or infection by blood transfusion. Epidemics of hepatitis E occur in Central and South East Asia, North and West Africa, and in Mexico. However, sporadic cases of hepatitis E are well documented in other regions, suggesting a global distribution. Hepatitis E is extremely common in developing countries, where it accounts for over 50% of cases of acute viral hepatitis.

Typical signs and symptoms of hepatitis include jaundice, anorexia, an enlarged, tender hepatomegaly, abdominal pain and tenderness, nausea and vomiting, and fever, but in most instances the disease is subclinical or causes a mild self-limiting illness often not requiring hospitalization. Diagnosis is based on serological tests for elevated antibody levels to hepatitis E based on reverse transcriptase PCR. At present, no commercially available vaccines exist for the prevention of hepatitis E. Occasionally, a fulminant form of hepatitis develops, with overall mortality rates ranging between 0.5% and 4.0%. Fulminant hepatitis occurs more frequently in pregnancy, usually in the third trimester, and is attended with a reported mortality rate of 20%.

Non-A, non-B hepatitis

This term is reserved for hepatitis by as yet unidentified viruses and the diagnosis is made by exclusion of the known hepatotropic viruses and CMV. Hepatitis C was grouped within this category before its isolation and cloning. Non-A, non-B (NANB) hepatitis (with unidentified viruses) remains the problem facing the transfusion of blood and blood products (as there are no tests to screen donors who may harbour these viruses). It accounts for the vast majority of all transfusion-related hepatitis and approximately 7–10% of transfused patients develop the infection. NANB is also transmitted by clotting factor concentrates (factors VIII and IX) and for this reason is very common in haemophiliac patients. Outbreaks following intravenous administration of immunoglobulin have been reported. The incubation period of NANB hepatitis varies widely but averages 8 weeks. The disease is clinically indistinguishable from hepatitis A and B and may indeed cause fulminant liver failure. However, in most instances, the hepatitis is mild and may only be represented by an elevation of the serum aminotransferase enzymes. Nonetheless, significant cohorts of patient who contract NANB infection go on to develop chronic liver disease. The clinical significance of the recently isolated hepatitis G virus reported in patients after blood transfusion remains uncertain.

There is increasing evidence for a non-A, non-B, non-C virus that causes aplastic anaemia and fulminant hepatitis. The affected patients develop jaundice followed, within 90 days, by aplastic anaemia. Some patients develop the aplastic anaemia after liver transplantation for NANB hepatitis.

Acquired immunodeficiency syndrome and related conditions

There is no doubt regarding the worldwide devastation caused by AIDS and its allied conditions: persistent generalized lymphadenopathy (PGL) and AIDS-related complex (ARC). Since the first few cases were recognized in the USA in 1981, estimates from the WHO from 180 countries show that over 10 million people are infected with the virus (HIV positive) but are currently well. The majority of these will, in time, develop the disease and eventually die from it. AIDS has particular relevance to surgeons and other healthcare workers who, by virtue of their occupation, are at risk, although with suitable precautions the risk of acquiring the infection from patients is small.

Virology and immune deficiency

The first virus was isolated in France from a patient with lymphadenopathy and was therefore called lymphadenopathy-associated virus. In the USA the virus was first isolated from AIDS patients and was thus named human T-cell lymphotropic virus type III (HTLV-III) to distinguish it from other known lymphotropic viruses (home in on lymphocytes) such as HTLV-I and HTLV-II. The former is associated with a special form of leukaemia/lymphoma. There is now international agreement that the term HIV should be used to describe the agent responsible for AIDS and allied conditions (PGL, ARC). More recently, a variant, HIV-2, has been isolated from African patients, but there does not appear to be any material difference in the disease potential between these two viruses and the collective term HIV is recommended for either agent.

The structure of HIV is shown in Figure 11.22. It is a retrovirus, i.e. its genome contains only RNA; in order to replicate once it invades the host cell, the viral RNA is transcribed into DNA by a special viral enzyme called reverse transcriptase, i.e. new viral RNA is made using the host DNA as the template (reverse of the normal process: DNA makes RNA makes protein). Once introduced into the bloodstream, HIV homes in on the T-helper lymphocytes, monocytes and macrophages that carry the surface receptor, CD4, on to which the virus latches. The T-helper lymphocytes (CD4) thus colonized are progressively destroyed. By contrast, the colonized macrophages, although providing a reservoir for continued viral replication, seem to be resistant to its destructive effects. HIV also colonizes the neural tissue and this accounts for the dementia that develops in some patients.

The T-suppressor/cytotoxic lymphocytes are not colonized because they contain a different receptor (CD8). The replication and expression of the virus in the T-helper lymphocyte terminates when the cell dies prematurely. The longitudinal consequence of this HIV/T-helper lymphocyte interaction for the infected patient are impaired function and progressive gradual depletion of the T-helper lymphocytes, reversal of the helper/inducer (CD4) to suppressor/cytotoxic (CD8) cell ratio and progressive decrease in the level of detectable virus in the blood. The immunological depression is characterized by several abnormal parameters (Box 11.1).

Laboratory evidence of human immunodeficiency virus infection

The current routine tests are based on the detection of antibodies to circulating HIV. Following infection, antibodies to the various antigens appear in the circulation. These do not appear to confer any benefit but are used in the diagnosis (HIV antibody testing). Regrettably, there is a long period (usually 6 weeks but sometimes several months) between infection and the appearance of antibodies. There is, therefore, a long 'window' when an individual may have negative antibody testing to HIV and still be infected.

Other tests include viral cultivation from the patient's lymphocytes, detection of circulating viral P24 antigen (useful in helping to establish infection in an infant whose HIV antibodies may be maternal) and the use of PCR, by which the very small amount of viral DNA in infected cells can be greatly amplified. The PCR method is the gold standard test for the confirmation of HIV infection.

Transmission of human immunodeficiency virus infection

The proven vehicles of infection are blood, semen and vaginal secretions. The disease is transmitted by penetrative unprotected sexual intercourse (homosexual and heterosexual), needle sharing by drug misusers and blood products. Prior to the introduction of HIV testing of donors, a large proportion of haemophiliacs were infected by contaminated clotting factor concentrates, as were a few patients who received blood transfusion to cover surgical procedures, especially cardiac operations. With the introduction of testing of blood donations, this route of transmission has now been eliminated but regrettably many of the infected haemophiliacs have died of AIDS and some have transmitted the disease to their sexual partners. The largest two groups contracting HIV infection are still active homosexuals (especially those with multiple partners) and drug addicts and users, in whom HIV is spread by sharing contaminated needles.

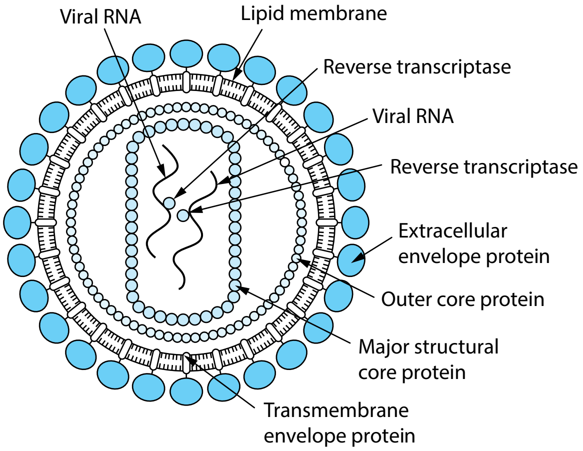


Figure 11.22 The structure of human immunodeficiency virus.

BOX 11.1 Parameters of immunodeficiency disorder in AIDS, persistent generalized lymphadenopathy (PGL) and AIDS-related complex (ARC)

- Decreased CD4+ lymphocytes and lymphopenia
- Decreased ratio CD4:CD8 lymphocytes
- Reduced cytotoxic response
- Reduced monocyte function
- Increased immunoglobulins
- Decreased blastogenic response of lymphocytes to mitogens
- Cutaneous anergy to multiple skin test antigens
- Increased levels of circulating immune complexes

Babies of HIV-positive mothers have a 25% risk of contracting the infection *in utero* or perinatally and by breast feeding. By contrast, parents of infected children have not been shown to be at risk. Oral zidovudine during late pregnancy and labour reduces the rate of mother to child transmission of HIV by 51%. The longer course reduces it by 67%. Breast feeding with the short remains a problem, as this is responsible for one-third of the maternal transmission. Thus, in short-course prevention, breast feeding has to be replaced by formula feeding of the babies.

There is no evidence that HIV is transmitted by non-sexual social contact (including sharing meals), accidental contamination with blood during participation in contact sport or by blood-sucking insects. Transmission by human bites is possible and occurs if blood is transmitted.

The risk of transmission to healthcare workers is due to certain mishaps:

- direct percutaneous inoculations of infected blood, e.g. accident by sharps (needle pricks, scalpel stab injuries, etc.)
- spillage of infected blood onto skin may introduce infection through minute scratches or abrasions
- contamination of mucosal surfaces by infected blood, e.g. accidental splashing of eyes
- transfer of infected material via environmental surfaces, e.g. blood-contaminated equipment and instruments.

The question concerning the risk involved in mouth-to-mouth resuscitation in an emergency remains unsettled, although there have been no such reported incidents to date.

Clinical features of human immunodeficiency virus infection

The majority of patients do not develop any symptoms after infection and are discovered to be HIV positive with testing. A small proportion of individuals acquire a self-limiting illness that resembles infectious mononucleosis a few weeks to months after the infection and others exhibit transient encephalitis. Several months later, some infected subjects develop a more lasting generalized enlargement of the lymph nodes, PGL. Significant symptoms start to appear with disease progression and these include weight loss, chronic diarrhoea, minor opportunistic infections and candidiasis. The deterioration in health and physique is clear-cut with progression to full-blown AIDS, the clinical features of which vary in accordance with the type of indicator illness that develops. The average incubation period from the time of infection to full-blown AIDS is 8 years, although faster progression is well documented. In the past, the available evidence suggested that eventually most HIV-positive subjects developed AIDS, although the time frame was variable This situation has changed drastically with highly active antiretroviral therphy (see Treatment). The clinical markers associated with an increased risk of progression to AIDS are constitutional symptoms, oral candidiasis and oral hairy leucoplakia. The laboratory indices that are also predictive of progression to AIDS include anaemia, lymphopenia, neutropenia, raised erythrocyte sedimentation rate (ESR), reduced T-helper (CD4) lymphocytes, raised CD8 lymphocytes, elevated viral antigenaemia (P24) and reduced anti-P24 antibodies.

The following clinical definitions are now accepted:

- AIDS: acquired immunodeficiency syndrome is defined as a syndrome in which a person has a reliably diagnosed disease that is at least moderately indicative of an underlying cellular immune deficiency (e.g. an opportunistic infection or Kaposi's sarcoma in a person younger than 60 years) but who, at the same time, has no known underlying cause of cellular immune deficiency or any other cause of reduced resistance reported to be associated with that disease. The reliably diagnosed disease is referred to as the indicator disease. The nature of this varies according to the absence (Box 11.2) or presence of laboratory evidence of HIV infection (Box 11.3). Absence of laboratory evidence may arise because the tests for HIV were not performed (for any reason) or gave inconclusive results. In the UK, the most commonly diagnosed indicator illness for AIDS is pneumonia due to *Pneumocystis carinii*.
- PGL: persistent generalized lymphadenopathy is defined as lymphadenopathy at two or more non-contiguous, non-inguinal sites for 3 months or longer in the absence of any current illness or medication known to cause enlarged lymph nodes.
- ARC: AIDS-related complex is defined as weight loss of more than 10%, intermittent fever of more than 38°C, intermittent or continuous diarrhoea, fatigue and malaise, night sweats and PGL.

In addition, both PGL and ARC must be accompanied by at least two abnormalities indicative of an immune deficiency/disorder. With the definitions and criteria shown in Boxes 11.1 and 11.2, an alternative useful clinical classification of HIV infection, which embraces all stages, is shown in Table 11.4.

Tumours in acquired immunodeficiency syndrome

The tumours that develop in AIDS patients are Kaposi's sarcoma, primary cerebral lymphoma and non-Hodgkin's lymphomas. Kaposi's sarcoma is one of the most common manifestations of AIDS and carries a high mortality (usually from severe opportunistic infections) with a median survival of 18–24 months. The tumour is a multifocal metastasizing malignant reticulosis, with features resembling those of angiosarcoma. It involves principally the skin (multiple red violaceous nodules), mucous membranes and lymph nodes, although visceral lesions are sometimes encountered. The disease in AIDS sufferers differs in several respects from Kaposi's sarcoma developing in immunocompetent individuals. There is now an established association of AIDS Kaposi's sarcoma with CMV infection. Significant palliation can be obtained by local radiotherapy.

The non-Hodgkin's lymphomas which develop in AIDS patients are of either -cell or indeterminate origin and of certain histological types: small non-cleaved lymphoma (either Burkitt's or non-Burkitt's), immunoblastic lymphoma, large-cell lymphoma, diffuse histiocytic lymphoma, diffuse undifferentiated lymphoma and high-grade lymphoma.

Antiviral therapy for acquired immunodeficiency syndrome

There have been considerable advances in antiretroviral drug therapy such that adequately treated AIDS with a combination of three or more anti-HIV drugs, sometimes referred to as highly active antiretroviral therapy (HAART), is compatible with long-term survival with the HIV virus.

The classes of antiretroviral drugs used in AIDS are:

1. Nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) interfere with the action of an HIV protein called reverse transcriptase, which the virus needs to make new copies of itself.
2. Non-nucleoside reverse transcriptase inhibitors (NNRTIs) also stop HIV from replicating within cells by inhibiting the reverse transcriptase protein.
3. Protease inhibitors (PIs) inhibit protease, which is another protein involved in the HIV replication process.

BOX 11.2 Indicator diseases for case definition of AIDS for surveillance purposes without laboratory evidence regarding HIV infection (Centers for Disease Control, Atlanta, GA)

- Candidiasis of the oesophagus, trachea, bronchi or lungs
- Extrapulmonary cryptococcosis
- Cryptosporidiosis with diarrhoea >1 month
- Cytomegalovirus disease of an organ other than liver, spleen or lymph node in a patient >1 month old
- Herpes simplex virus infection causing a mucocutaneous ulcer that persists longer than 1 month; or bronchitis, pneumonitis or oesophagitis for any duration affecting a patient >1 month old
- Kaposi's sarcoma affecting a patient <60 years old
- Primary lymphoma of the brain affecting a patient <60 years old
- Lymphoid interstitial pneumonia and/or pulmonary lymphoid hyperplasia affecting a child <13 years old
- *Mycobacterium avium* complex or *Mycobacterium kansasii* disease, disseminated (at a site other than or in addition to lungs, skin, cervical or hilar lymph nodes)
- *Pneumocystis carinii* pneumonia
- Progressive multifocal leucoencephalopathy
- Toxoplasmosis of the brain affecting a patient >1 month old

4 Fusion or entry inhibitors prevent HIV from binding to or entering human immune cells.
5 Integrase inhibitors interfere with the integrase enzyme, which HIV needs to insert its genetic material into human cells.

BOX 11.3 Indicator diseases for case definition of AIDS for surveillance purposes with laboratory evidence regarding HIV infection (Centers for Disease Control, Atlanta, Georgia)

- Bacterial infections, multiple or recurrent (any combination of at least two within a 2 year period), of the following types affecting a child <13 years old: septicaemia, pneumonia, meningitis, bone or joint infection, or abscess of an internal organ or body cavity (excluding otitis media or superficial skin or mucosal abscesses), caused by *Haemophilus*, *Streptococcus* or other pyogenic bacteria
- Coccidioidomycosis, disseminated (at a site other than or in addition to lungs or cervical or hilar lymph nodes)
- HIV encephalopathy (HIV/AIDS dementia)
- Histoplasmosis, disseminated (at a site other than or in addition to lungs or cervical or hilar lymph nodes)
- Isosporiasis with diarrhoea persisting >1 month
- Kaposi's sarcoma at any age
- Primary lymphoma of the brain at any age
- Other non-Hodgkin's lymphoma of -cell or unknown immunological phenotype
- Any mycobacterial disease caused by mycobacteria other than *M. tuberculosis*, disseminated (at a site other than or in addition to lungs, skin, or cervical or hilar lymph nodes)
- Disease caused by *M. tuberculosis*, extrapulmonary (involving at least one site outside the lung, regardless of whether there is a concurrent pulmonary involvement)
- *Salmonella* (non-typhoid) septicaemia, recurrent
- HIV wasting syndrome (emaciation, slim disease)

Table 11.4 Alternative useful clinical classification of HIV infection

Group	Definitio
I	Acute infection: infectious mononucleosis-type illness
II	Asymptomatic infection: asymptomatic HIV-positive subjects
III	Persistent generalized lymphadenopathy (PGL)
IV	
A	Constitutional disease: fever, diarrhoea and weight loss lasting >1 month
B	Neurological disease: dementia, myelopathy and peripheral neuropathy
C	Secondary infectious disease: opportunistic
C_1	Specified in US Centers for Disease Control surveillance definition of AIDS
C_2	Other infections: oral hairy leucoplakia, nocardiosis, oral candidiasis, etc.
D	Secondary cancers: Kaposi's sarcoma, non-Hodgkin's lymphomas, primary cerebral lymphoma
E	Other conditions: other disorders attributable to HIV infection

The currently used drugs in these five categories are:

1 NRTIs: lamivudine, abacavir, zidovudine, didanosine, emtricitabine.
2 NNRTIs: delavirdine, efavirenz, etravirine, nevirapine, rilpivirine, atazanavir.
3 PIs: amprenavir, fosamprenavir, atazanavir, darunavir, indinavir, ritonavir, saquinavir, maraviroc.
4 Fusion or entry inhibitors: enfuvirtide.
5 Integrase inhibitors: raltegravir.

Protease inhibitors are generally less suitable for starting treatment in resource-limited settings owing to the cost, number of pills which need to be taken, and the particular side effects caused by protease drugs.

The CD4 test

The CD4 test is used to determine the need for treatment since it measures the number of T-helper cells (CD4 count), which normally ranges between 500 and 1200 cells/mm^3. HIV drug treatment is indicated when the CD4 count is <350 cells/mm^3. The WHO 2010 guidelines recommend starting treatment for all patients with CD4 counts of <350 cells/mm^3 in all countries but the treatment guidelines may differ from WHO recommendations, e.g. in the USA, where treatment is recommended for patients with a CD4 count between 350 and 500 cells/mm^3.

A basic clinical assessment should be carried out before treatment to establish the general condition and body weight of the patient, existing disease (hepatitis, tuberculosis), injection drug use, pregnancy in females, psychiatric illness and any oral medications which might cross-react with the antiviral drugs. The patients must be psychologically ready to start and persist with therapy despite the demanding regimen and associated side effects. HAART first-line therapy consists of a combination of three or four drugs: a typical antiretroviral combination consists of two drugs from the NRTI class and one drug from another class.

If facilities are available preliminary tests for HIV drug resistance are carried out before finalizing the regimen in the individual patient as some HIV strains have become resistant to a certain class of drug. Reports indicate that approximately one in 10 HIV-infected individuals in Europe and the USA are infected with a type of HIV that is resistant to at least one class of antiretroviral drug. Testing for drug-resistant strains may not, however, be available in some poor countries. Another practical consideration related to interaction with other medications. This is a common problem among older patients with HIV, as they are usually on drugs for age-related illnesses. Some drugs may cause side effects or decrease the effectiveness of some antiretroviral drugs.

Once patients start HIV treatment, it is essential that (1) they comply fully (adhere) with the treatment and (2) they are monitored regularly to ensure that the treatment is effective in reducing the viral load and improving the CD4 count. Adherence signifies that the patient is taking the drugs exactly as prescribed, i.e. taking all the drugs at the right time and in accordance with instructions. Anything below full adherence has been associated with increased viral load and the development of drug resistance.

The aim of antiretroviral treatment is to keep the viral load as low as possible. The viral load test is carried out shortly after initiation of antiretroviral treatment. Successful therapy is signified by a dramatic drop to an undetectable level (below 50 copies/mL) within 3–6 months. However, some patients never reach undetectable levels. Viral load tests are then carried out every few months.

Side effects

Side effects of antiretroviral drug therapy are common. In some people these are mild, but in others they are severe enough to warrant use of alternative drugs. Another complication is immune reconstitution inflammatory syndrome (IRIS), which occurs in a minority of patients soon after treatment is started. It is due to an excessive response by the recovering immune system to opportunistic infections already present, but which were previously dormant. Symptoms of IRIS are often mild, but occasionally they can be life-threatening. If the symptoms of IRIS are mild, no change in therapy is needed as the symptoms disappear within a few weeks. However, with severe IRIS involving severe opportunistic infections, e.g. cryptococcal meningitis or tuberculosis, it may be necessary to stop antiretroviral therapy until the infection is treated.

A change of treatment is indicated if the antiretroviral regimen fails to slow down the viral replication. Some experts recommend changing as soon as the viral load starts to rise, whereas others recommend monitoring the trend of the viral load before making a decision to change the regimen. When viral load testing is not available, the WHO staging system for HIV disease (determines the stage of HIV infection based on clinical symptoms) is used instead:

Stage I:	HIV disease is asymptomatic and not categorized as AIDS.
Stage II:	includes minor mucocutaneous manifestations and recurrent upper respiratory tract infections.
Stage III:	includes unexplained chronic diarrhoea for longer than a month, severe bacterial infections and pulmonary tuberculosis.
Stage IV:	includes toxoplasmosis of the brain, candidiasis of the oesophagus, trachea, bronchi or lungs and Kaposi's sarcoma; these diseases are used as indicators of AIDS.

Salvage therapy

Salvage therapy is the term often used to describe the treatment for those who are resistant to drugs in the three original drug classes: NRTIs, NNRTIs and PIs. In these patients it may be difficult to find a drug regimen that suppresses the viral load to undetectable levels. The option is the use of much higher doses. The two additional classes of drugs introduced since 2003 (fusion or entry inhibitors and integrase inhibitors) are better treatment options in these resistant cases. It should be stressed that although antiretroviral drugs suppress HIV they do not eliminate the risk of HIV transmission completely, even when the viral load becomes undetectable. Thus, unprotected sex between two HIV-positive people is not without risk in view of the many different strains of HIV. For the majority of patients, HAART results in a near normal life expectancy.

Acquired immunodeficiency syndrome and hospital surgical practice

Experience in the UK and other countries suggests that the risk to healthcare workers is low provided certain universal and specific precautions are adopted routinely. As many infected individuals are unaware that they are HIV positive, all patients should be assumed to be infectious for HIV and other blood-borne pathogens. It is unsafe to apply universal precautions only in high-risk groups and in patients known to be HIV positive. The universal precautions (Box 11.4) apply to all health workers exposed to the following body fluids: blood, amniotic fluid, pericardial fluid, peritoneal fluid, pleural fluid, synovial fluid, cerebrospinal fluid, and semen and vaginal secretions. Universal precautions do not apply to exposure to faeces, nasal secretions, sputum, sweat, tears, urine and vomit. Although saliva is not regarded as a risk vehicle, precautions are needed in dental and oral work, as the saliva is then likely to be contaminated with the patient's blood.

When blood samples are obtained the doctor should wear disposable gloves. Plastic aprons and eye protection are needed when splashing or spurting is likely. Blood should be collected preferably with Vacutainers and on no account must the needle be resheathed. In high-risk cases, all specimens sent to the laboratory must be sealed in a plastic bag and marked with a red hazard sticker, as for hepatitis B. Accidental spillage of blood should be treated with 1% sodium hypochlorite solution. The specific guidelines for operating theatre staff are outlined in Box 11.5. All staff involved in the operating theatre should wear plastic aprons, disposable gowns, goggles, gloves and impermeable leg and footwear.

Action in the event of an accident involving possible transmission to medical and nursing staff

The rate of transmission resulting from accidental inoculation of healthcare workers with blood from individuals known to be infected has been reported to be of the order of 0.13–0.55, i.e. 1:770 to 1:200 chance of infection as a result of any single event. When such an accident occurs, the following procedure is necessary:

- Contaminated areas (spillage and splashes) or injury site (from sharps) should be immediately and thoroughly washed with soap and water.
- The incident is reported to the occupation health officer.
- If the HIV status of the source patient is not known, the patient consent's for HIV testing is obtained. This is usually forthcoming, but, if the patient declines consent, the statement by the General Medical Council becomes relevant: 'Only in the most exceptional circumstances, where a test is imperative in order to secure the safety of persons other than the patient, and where it is not possible for the prior consent of the patient to be obtained, can testing without explicit consent be justified.' In such a difficult situation, this statement implies that HIV testing of a blood specimen that had previously been obtained from the patient for other purposes is permissible.
- The exposed health worker should be counselled and expert advice sought. Consent for HIV testing from the individual concerned should be obtained, in which event, an immediate specimen (baseline) followed by others at intervals are necessary. Seroconversion usually occurs within 3 months.

BOX 11.4 Universal precautions against the spread of AIDS

- These apply to all patients irrespective of risk category or HIV status (known or unknown)
- Care in the handling of sharps: needles, scalpels, sharp instruments, etc.
- All cuts and abrasions on patients and staff should be covered with a waterproof dressing. Personnel should wear plastic aprons and disposable gloves when dealing with blood or secretions
- Parenteral procedures should be kept to a minimum
- External surfaces of equipment and bench surfaces which may have been contaminated by blood and other secretions should be wiped with a fresh preparation of sodium hypochlorite solution 1% (10 000 p.p.m. available chlorine) or 2% activated glutaraldehyde. Contaminated gloves, paper tissues and cotton wool should be incinerated
- In the event of the death of a person with AIDS, the body should be wrapped in impermeable plastic sheeting or a polythene cadaver bag before coffining
- Equipment being sent for maintenance and servicing should be disinfected with glutaraldehyde before leaving the ward or theatre
- Disposable equipment should be used wherever possible, and reusable equipment immersed in 2% glutaraldehyde for 1 hour before being returned for processing
- Walls and floor should be cleaned with soap and water

Fungal infections of surgical importance

From the clinical standpoint, fungal infections are best divided into cutaneous, subcutaneous and deep. Risk factors include prolonged antibiotic therapy that encourages growth of commensal fungal organisms, and leucopenia and T-helper lymphocyte (CD4) depression that are associated with invasive and potentially fatal fungal disease. The *cutaneous* (superficial) mycoses have a worldwide distribution and include dermatophytosis (ringworm), superficial candidosis (thrush) and pityriasis versicolor. The *subcutaneous* fungal infections are largely restricted to the tropical and subtropical regions, and often arise by direct implantation of soil fungal organisms through the skin of the soles of the feet, e.g. mycetoma and sporotrichosis.

A number of conditions result from *deep-seated* fungal infections that may be difficult to treat and prove fatal, particularly in immunosuppressed or debilitated patients. These include some fungal infections that have similar clinical characteristics, e.g. histoplasmosis, cryptococcosis, coccidiomycosis and blastomycosis. They are all caused by inhalation of infected soil dust and the primary lesion is pulmonary. Although usually self-limiting, the disease may spread via the bloodstream to involve other organs, causing serious or fatal illness. Many of the histological features of these deep-seated fungal infections are similar to those found in tuberculosis.

BOX 11.5 Guidelines for operating theatre staff

- It is justifiable to set aside a theatre and staff solely for HIV-positive patients
- Shaving should be avoided
- The anaesthetic room staff should wear disposable masks, plastic aprons, gloves and overshoes
- The anaesthetic machine and work surfaces should be stripped of all but essential equipment
- All theatre staff should wear impermeable overshoes, which must not be taken out of the theatre
- All staff, scrubbed and unscrubbed, should wear disposable gowns, plastic aprons, goggles and gloves (double for the surgeons and scrub nurse)
- Disposable drapes should be used. Swabs should be counted on a polythene sheet on the floor, not on the swab rack
- Only disposable scalpels are used. All instruments should be handed to and from the surgeon on a tray such that the surgeon or nurse picks the instrument without any direct transfer from nurse to surgeon and vice versa
- Suction bottles should be half filled with freshly prepared 2% glutaraldehyde solution
- Spilt blood or body fluids must be diluted with fresh 2% glutaraldehyde and mopped up with white paper towels. All consumables should be placed in a watertight bag for plastic incineration
- Splashes of blood or body fluids or accidental puncture wounds should be immediately and thoroughly washed with soap and water and the consultant should then inform the occupational health service of such an incident
- The operating nurse must ensure that the patient's skin is completely free of blood after the operation
- The patient should wear a clean operating gown before transfer back to the ward

Finally, *systemic mycoses* may arise from superficial and deep-seated fungal infections when the patient's immune response is depressed or the patient is critically ill. These are commonly encountered in ICU patients.

Candidosis (candidiasis)

The *Candida* species of fungi occur as commensals in the mouth, alimentary tract and vagina. Infection with *C. albicans* is common in clinical practice, and usually arises as an opportunistic superinfection in debilitated or critically ill and immunosuppressed patients.

The infection is encountered in several clinical forms. Oral candidosis (thrush) and the vaginal and perianal infections are common and usually occur as a consequence of antibiotic therapy (Figure 11.23). Vaginal candidosis is also found in conditions that result in a low pH of the vaginal secretions from excess glycogen in the vaginal cells, e.g. pregnancy, diabetes and females on the contraceptive pill. Cutaneous circumoral infection (perlèche) and chronic paronychia may also occur in these patients.

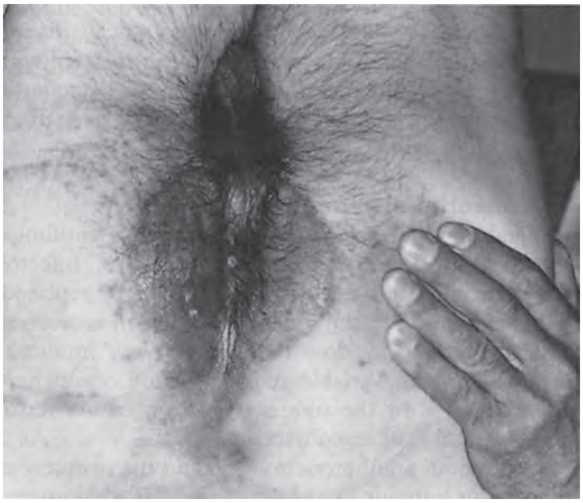


Figure 11.23 Perianal candidosis.

Alimentary and pulmonary candidosis

The oral infection may extend to involve the gastrointestinal tract, particularly the oesophagus, following prolonged therapy with broad-spectrum antibiotics. Both alimentary and pulmonary candidosis may arise as opportunistic infections in the presence of impaired cell-mediated immunity, e.g. immunosuppression from illness, malignancy or drugs.

Generalized (systemic) candidosis

This is the most serious form and is characterized by bloodstream spread to any organ, but most commonly to the lungs, kidneys and the endocardium. This serious systemic infection is encountered in patients on parenteral nutrition, patients with prosthetic devices, transplant patients and critically ill patients. The general clinical features of systemic candidosis include high fever, hypotension, rigors and renal impairment. The reported overall survival following appropriate systemic antifungal therapy is 65–70%.

Pulmonary candidosis

There are two types of pulmonary candidosis: monilial bronchitis and monilial pneumonia. The treatment is with 5-flucytosine, which is the drug of choice if the organism is sensitive; otherwise, intravenous or inhaled amphotericin B may be tried (see below).

Monilial bronchitis

This occurs in infancy, fibrocystic lung disease and debilitated elderly patients. The bronchial walls are studded with greyish-yellow plaques containing fibrin, hyphae and spores. Clinically, there is an irritating cough productive of scanty, 'milky' sputum. The chest radiograph may be normal or show basal striations.

Monilial pneumonia

This is a necrotic pneumonia that occurs in critically ill and debilitated patients, giving patchy, ill-defined shadows on the chest radiograph. The patient has a high fever with tachycardia, dyspnoea and cough productive of blood-stained sputum.

Chronic mucocutaneous candidosis

Several types of primary immunological disease affecting the T-lymphocyte function are associated with widespread development of chronic mucocutaneous candidosis. The immunological deficit varies from patient to patient.

Treatment of systemic fungal infections

General considerations are important since antifungal therapy alone may be insufficient. Thus, infected prosthetic implants need to be removed or replaced. Whenever possible, the immune depression is reversed at least partially by dose reduction of any immunosuppressive drug. Variable and conflicting results have been obtained by the administration of transfer factor and transfusion of leucocytes.

Intravenous amphotericin B remains the mainstay of therapy for systemic fungal infections. It is administered in 5% dextrose, initially in a dose of 0.25 mg/kg/24 h and increasingly gradually over a period of 4–6 days to 1.0 mg/kg/day. The daily dose is administered over a period of 3–4 hours. Since it deteriorates rapidly in solution on exposure to sunlight, the intravenous solution must be shielded. Treatment is monitored by blood levels, which should be kept in the 1–3 μg/mL range. The main disadvantage of amphotericin B is its toxicity: rigors, hypotension, phlebitis, hepatic damage, renal damage, anaemia, leucopenia and hypokalaemia. The prior intravenous administration of either hydrocortisone (50–100 mg) or chlorpheniramine (10 mg) is said to reduce some of the early side effects. In order to minimize toxicity, amphotericin B is often administered in a reduced dose in combination with other drugs. Various combinations are used, including amphotericin B + 5-flucytosine for cryptococcosis and rifampicin + amphotericin B for candidosis and blastomycosis. Altering the rate of infusion and alternate-day treatment may help to reduce the incidence and severity of side effects. Other measures that are used to reduce toxicity include heparin infusion against phlebitis, and mannitol and bicarbonate to minimize renal damage. 5-Flucytosine is a synthetic antifungal agent that can be administered orally or intravenously. It is active mainly against *Candida* and *Cryptococcus.* Apart from its side effects (diarrhoea, rashes, leucopenia and hepatitis), its main disadvantage is the rapid emergence of resistance.

A number of imidazoles have significant antifungal activity. The two most often used are ketoconazole, which is administered orally (200–600 mg), and miconazole, which is given intravenously since it is poorly absorbed after oral administration. Ketoconazole has fewer side effects than miconazole, which may cause rashes, phlebitis, ventricular tachycardia and anaphylaxis. Drug resistance to the imidazoles is rare and they have a broad spectrum of activity against many fungi (except for aspergilli) and some Gram-positive bacteria. All other imidazoles (clotrimazole, econazole, tioconazole) are mainly used as topical antifungal agents. Oral ketoconazole is advocated as prophylaxis in transplant patients.

Treatment of cutaneous fungal infections

Topical polyenes (natamycin, nystatin, candicidin) are used as 1–2% creams or ointments, as suspensions or tablets (oral or vaginal). The topical imidazole preparations are also effective. In chronic mucocutaneous candidosis associated with T-lymphocyte immune defects, treatment with transfer factor gives varying results.

Mycetoma

This is a localized infection of the skin and subcutaneous tissues of the extremities, usually of the feet and less frequently of the hands. It is found in developing countries. The infection is caused by organisms normally resident in the soil, which are implanted by thorn and similar injuries in individuals who walk barefooted. *Bacterial mycetoma* is caused by infection with species of *Nocardia* or *Actinomyces* (actinomycetoma). *Fungal mycetoma* (eumycetoma, maduromycosis) is endemic in certain countries, e.g. India, and is caused by various fungi including *Madurella mycetoma, Allescheria boydii* and *Aspergillus nidulans.* The disease results in chronic suppuration with severe tissue destruction and multiple discharging sinuses (Figure 11.24). In both bacterial and fungal mycetoma, osteomyelitis of the bones of the foot may develop.

Aetiology

The most common fungi to cause mycetoma with black or red grains are *Leptosphaeria senegalensis* (Africa), *Madurella grisea* (Africa, South and Central America), *Madurella mycetoma* (worldwide) and *Pyrenochaeta romeroi* (Africa, South America). Other fungal infections cause mycetoma with white grains, e.g. *Acremonium* species (Africa, Middle East), *A. nidulans* (Africa, Middle East), *Noetestudina rosatii* (Africa) and *Pseudallescheria boydii* (worldwide).

The common actinomycetes causing mycetoma usually with white/yellow grains are: *Actinomadura madurae* (worldwide), *Nocardia asteroides* (worldwide), *Nocardia brasiliensis* (Central America). *Actinomadura pelletieri* (Africa) and *Streptomcyes somaliensis* (North Africa, Middle East) cause mycetoma with brown or red grains in the discharge.

Clinical features

Mycetoma is more common in males than in females, and affects usually individuals aged 20–50 years. It presents as a subcutaneous single hard painless lump most commonly of the foot (70%), which persists and become indurated as it gradually involves the underlying muscles and bones. Eventually the lesion ulcerates through the skin with the formation of sinuses discharging pus containing the characteristic grains. The condition causes considerable deformity, rendering walking difficult, but does not cause pain, although itching and burning sensation are common, as is secondary bacterial infection including osteomyelitis and atypical mycobacterium infection.

The diagnosis of mycetoma is made on identifying grains and culture of pus. The colour of the grains often suggests the likely diagnosis: black or red grains in fungal infection, white grains in *Nocardia*, etc. Microscopy of smears stained with potassium hydroxide confirms the diagnosis and type of mycetoma, but Agar plate cultures at 25–30°C for up to 6 weeks may be needed for definite diagnosis.

Treatment

In bacterial mycetoma, treatment with the appropriate antibiotics together with surgical drainage and debridement may prove efficacious. Bacterial actinomycetoma responds well to treatment with prolonged appropriate antibiotic therapy (6–12 months). Single or combination treatment may be used: streptomycin injections; oral cotrimoxazole, amikacin, dapsone, rifampicin or minocycline. Surgical treatment is used for debridement with foot preservation, but severe disease with bone involvement may require amputation.

The fungal eumycetomas, by contrast, seldom respond to antifungal chemotherapy as the organisms acquire a protective cement sheath or develop thickened cell walls. Nonetheless, a trial of treatment with an antifungal agent, e.g. griseofulvin, 5-flucytosine or ketoconazole or itraconazole, may be effective in patients with early and limited disease, and without bone involvement. In most instances, however, amputation is necessary to eradicate the disease.

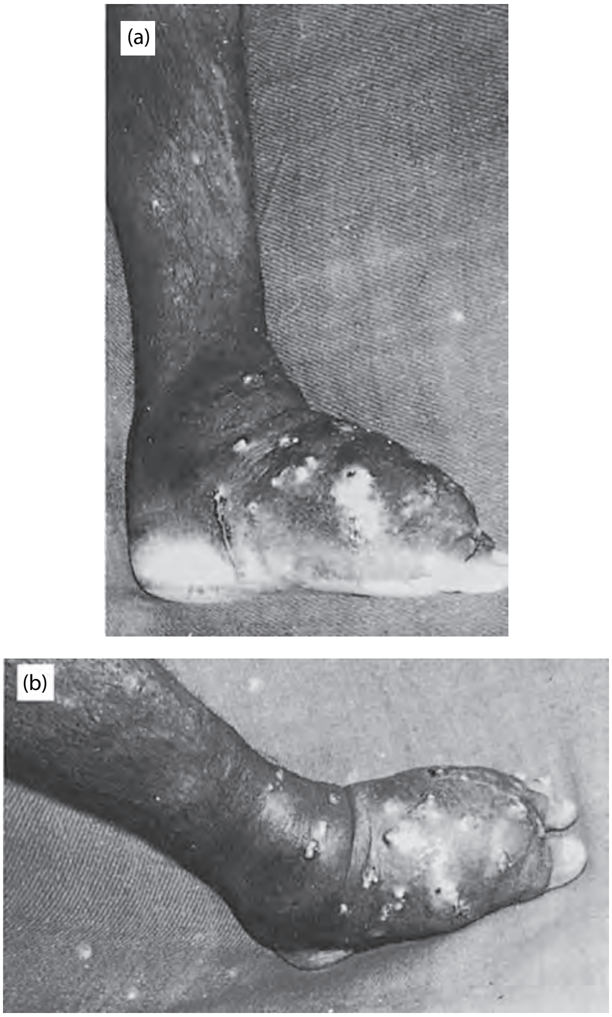


Figure 11.24 (a, b) Fungal mycetoma of the foot showing swelling, chronic induration with multiple sinuses and fungal granules. (Courtesy of Dr D.M. Muthukumarasamy, Kilpauk Medical College, India.)

Parasitic infestations of surgical importance

Hydatid disease

The disease can be caused by one of three species of the genus *Echinococcus: Echinococcus granulosus, Echinococcus multilocularis* and *Echinococcus oligettas.* Since the epidemiological and pathological features of these three tapeworms are very similar, a detailed description of *E. granulosus* only is given here. Hydatid disease, which is caused by the larval form of *E. granulosus,* has a cosmopolitan distribution, being particularly prevalent in sheep- and cattle-raising areas.

Parasitology

The adult *Echinococcus* is a small tapeworm that inhabits the upper part of the small intestine of canines, especially dogs and wolves. When the ova, which are passed in the faeces, are swallowed by humans or other intermediate hosts, e.g. sheep and cattle, the enclosed embryo is liberated in the duodenum. It penetrates the intestinal mucosa to reach the portal circulation, and is usually held up in the liver to develop into a hydatid cyst. If the embryo passes through the liver filter, it enters the general circulation and thus reaches the lungs and other parts of the body. Two main varieties of cyst occur: unilocular and multilocular. The unilocular hydatid cyst (Figure 11.25) develops a wall with two layers: an outer, thick, laminated layer and an inner (germinal) layer, which is composed of a protoplasmic matrix containing many nuclei. Around the cyst there is a connective tissue capsule formed by the tissues of the host. Bulb-like processes known as brood capsules arise from the germinal layer. The brood capsules undergo localized proliferation and invagination of their walls to form scolices (tapeworm heads). Each scolex has suckers and hooklets. Some of the brood capsules separate from the walls and settle to the bottom of the cyst as fine granular sediment 'hydatid sand'. As the hydatid cyst enlarges, invaginations of the wall may give rise to daughter cysts. In some cases in which no effective encapsulation occurs, the daughter cysts develop as a result of evaginations of the cyst wall, producing the multilocular or alveolar hydatid cysts. This variety of cyst is caused by *E. multilocularis.*

When the hydatid is eaten by definitive hosts, e.g. dogs, foxes or wolves, the numerous larvae develop into sexually mature worms in a few weeks. Dogs are usually affected when they eat the infected viscera of sheep or cattle (Figure 11.26). Humans acquire hydatid disease when they swallow infected ova as a result of their close association with dogs. Although infection is usually acquired in childhood, clinical symptoms do not appear until adult life. The dog faeces contaminating the fleeces of sheep can also be an indirect source of human infection.

Pathology

When the ingested ovum reaches the duodenum, the hexacanth embryo is released and penetrates the intestinal mucosa to reach the portal circulation and thence the liver, most commonly the right lobe. Hydatid cysts are usually single but may be multiple and involve the left lobe. Larvae may pass through the liver to infect other organs, e.g. lungs (71%), muscle (5%), brain (5%), spleen (2.5%), kidneys (2%), other abdominal organs (5%), bone (0.5%) and rarely the heart. During the stage of migration, there may be a mild reaction with fever and urticarial skin rash. There is an eosinophilic reaction to the larva itself, which subsides as the host's fibrous tissue capsule thickens. In humans the hydatid cyst may be the classic unilocular, the osseous or the alveolar.

The alveolar multiloculated hydatid cyst is usually found in the liver and has a sponge-like appearance (Figure 11.27). In the brain it resembles a bunch of grapes (Figure 11.28). The alveolar multiloculated cyst may metastasize.

The classic unilocular hydatid cyst may be sterile but is usually fertile and surrounded by a fibrous capsule. Rupture of fertile cysts, depending on their site, causes dissemination of daughter cysts with the formation of secondary metastatic hydatids. The escape of the cyst fluid and 'hydatid sand' may

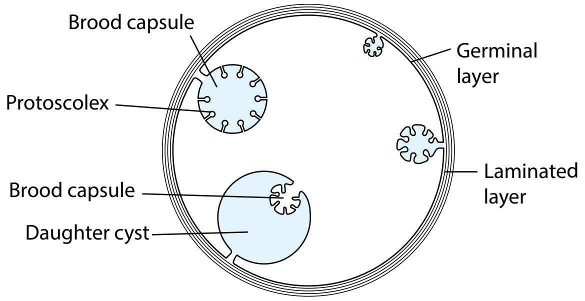


Figure 11.25 Composition of a unilocular hydatid cyst.

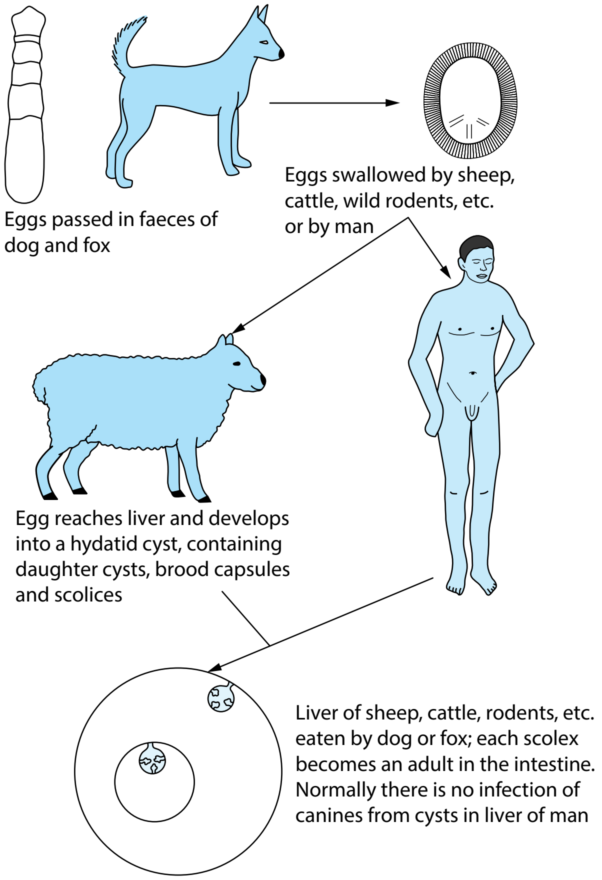

Eggs passed in faeces of dog and fox
Eggs swallowed by sheep, cattle, wild rodents, etc. or by man
Egg reaches liver and develops into a hydatid cyst, containing daughter cysts, brood capsules and scolices
Liver of sheep, cattle, rodents, etc. eaten by dog or fox; each scolex becomes an adult in the intestine. Normally there is no infection of canines from cysts in liver of man

Figure 11.26 Life cycle of *Echinococcus granulosus.*

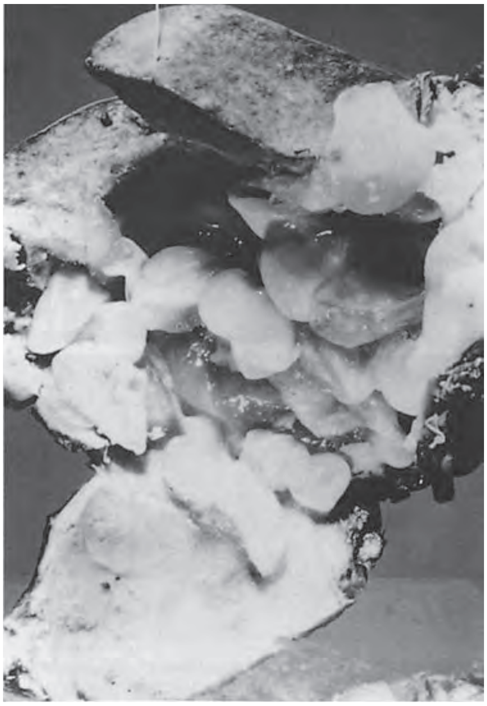


Figure 11.27 Hydatid disease of the liver.

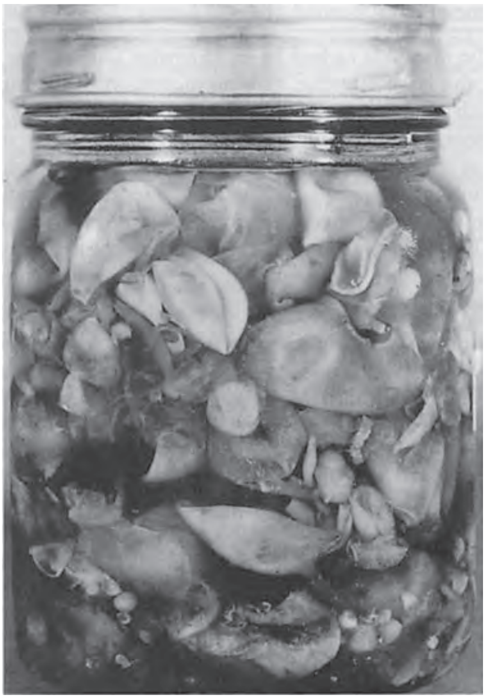


Figure 11.28 Hydatid cysts removed from the brain at necropsy.

cause a severe allergic reaction with urticarial lesions, pruritus, fever, abdominal pain, dyspnoea, cyanosis, delirium and syncope. In addition, there is marked eosinophilia. In bone the usual fibrous capsule of the host is not formed and, instead of being round, the cyst assumes an irregular branching shape as it penetrates the bony canals. Erosion of bone occurs and the medullary cavity is eventually invaded, when the cyst assumes its normal spherical form. The more highly vascularized areas, the epiphyses of long bones and the centres of the vertebral bodies, ilium and ribs, are the most frequently affected sites. Radiologically, they appear as rounded areas of rarefaction. Spontaneous fractures may occur.

Clinical features

The liver and lungs are the organs involved in the disease in over 90% of cases and hepatic hydatid disease accounts for 75% of these. Although the infection is usually acquired in childhood, a hepatic hydatid cyst may not produce symptoms until adult life and the interval between infection and manifestation of symptoms may exceed 30 years. At presentation the majority of cysts occur in the right lobe. Exogenous cysts may communicate with the biliary tract and these contain bile-stained fluid (Figure 11.29). Secondary infection may occur with *Salmonella* and pyogenic organisms.

Rupture of the cyst may result from secondary infection or trauma and may be accompanied by severe allergic manifestations that can be life-threatening. Rupture may involve the peritoneal cavity, the gallbladder, biliary tract, pleural cavity or hepatic veins, when secondary metastatic cysts develop in the lungs.

The alveolar cyst may occur in the liver but is most frequently found in the lungs in children. Infection of the cyst results in a chronic pulmonary abscess. The cyst may rupture as a result of coughing or secondary infection. The cyst may rupture into the bronchus, when scolices may be detected in the sputum, into the mediastinum or pleural cavity, with secondary dissemination and the development of pneumothorax and haemothorax. In the brain, the cysts are usually single and present with symptoms of a space-occupying lesion. Rupture may disseminate the lesions in the subarachnoid space of the spinal cord, which may also be involved by extension of a hydatid cyst involving the vertebrae or paravertebral tissues. Cerebral emboli, secondaries have been reported from primary hydatid cysts in the lung, liver and heart.

Hydatid disease of the heart may be first discovered at autopsy. During life, the cysts may rupture into the pericardium or into the chambers of the heart, with consequent pulmonary metastases or systemic embolization causing thrombosis, infarction or aneurysm formation. Constrictive pericarditis may complicate rupture into the pericardium.

Occasional involvement of the spleen and kidneys by solitary cysts is well documented. In the kidney rupture into the renal pelvis results in renal colic and dysuria, with passage of hydatid material in the urine. Rarely, there is involvement of other sites, e.g. orbit, ovaries, broad ligament and uterus.

Investigations

Moderate eosinophilia (300–2000/mm^3) is usually present and the immunoglobulin level is elevated. Hydatid contents (hooklets, scolices) may be detected in the faeces, sputum or urine if the cyst ruptures. A plain radiograph of the chest and abdomen may be suggestive (Figure 11.30), although more information is obtained by ultrasound examination and especially by CT, which gives valuable information on the contents, size and precise location of the cyst (Figure 11.31). CT is the best method of establishing the diagnosis of liver, pulmonary and cerebral hydatid disease.

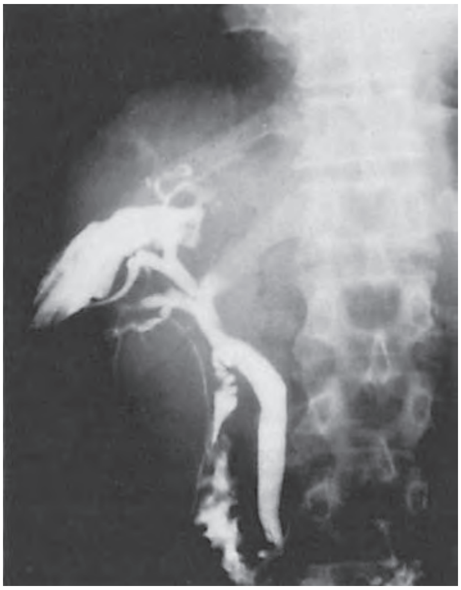


Figure 11.29 Hydatid cyst of the liver communicating with the right biliary tract.

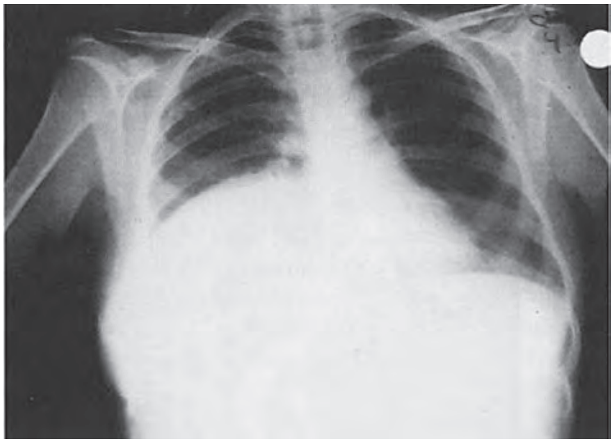


Figure 11.30 Posteroanterior chest radiograph outlining a large liver with marked elevation of the diaphragm caused by a hydatid cyst of the right lobe.

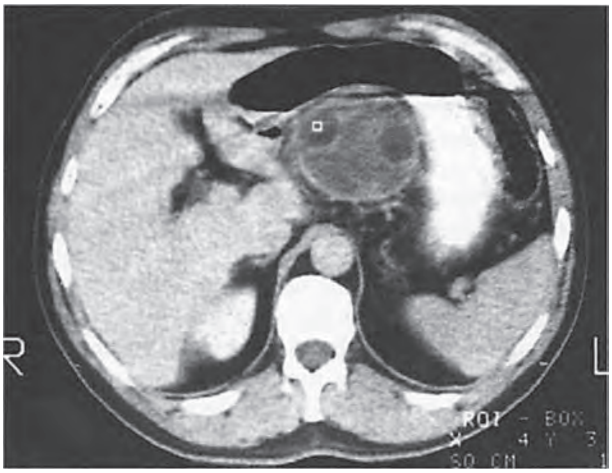


Figure 11.31 CT scan showing a large hydatid cyst of the pancreas containing daughter cysts. The pancreas is a rare site of involvement by hydatid disease. (Courtesy of Dr Moh'd Saad, Adan Hospital, Kuwait.)

The Casoni test using crude sterile hydatid fluid is unreliable. An intradermal test antigen made from an extract of lyophilized material of *E. multilocularis* from experimental infections in gerbils is preferable. The detection of circulating scolex antigen by countercurrent immunoelectrophoresis or the enzyme-linked immunosorbent assay (ELISA) gives the most reliable results. The dot-ELISA test is a field test for hydatid disease that can be carried out rapidly without specialized equipment. It detects antibody to purified parasite antigen (bound to strips of nitrocellulose paper) in a small finger-prick sample.

Treatment

The methods used to control the disease are aimed at prevention of access to infected carcasses by dogs, and registration and regular treatment of dogs with effective anthelmintics

The use of mebendazole often results in death and shrinkage of the hydatid disease, thereby avoiding the need for surgical intervention. Serum levels >1100 ng/mL 1–3 hours after an oral dose may be required to kill the parasite. To achieve this, a dose of 200 mg/kg/day of mebendazole in three divided doses is required for about 16 weeks. Albendazole, a more absorbable derivative, is even more effective. It is administered in a dose of 12–15 mg/kg/day for 28 days.

Surgical treatment

This has to be individualized to the pathology in the individual patient and the clinical presentation. The surgical approach may be open or laparoscopic, although most advise that the laparoscopic approach should be restricted to cysts in an accessible location, namely the left lateral and right anterior segments. The various procedures performed for hepatic hydatid cysts may be classified as (1) conservative or (2) radical. The conservative operations are:

- de-roofing of the cyst and extraction of the parasite with omentoplasty and external drainage of the cyst cavity: Mabit procedure
- de-roofing of the cyst with captionage of the cavity without drainage: Posadas procedure
- partial pericystectomy: leaving a deep part of the cyst within the liver
- marsupialization
- internal cystojejunal drainage
- external drainage: not recommended as it is associated with a high incidence of infective complications (30%) and biliary fistulae (8%) and has largely been replaced by omentoplasty, which carries a significantly lower incidence of biliary fistula and infection.

The more radical procedures are:

- pericystectomy (removal of cyst and 1–2 cm of normal parenchyma)
- total cystectomy (complete removal of cyst)
- liver resection.

Pericystectomy consists of removing the cyst together with a 1–2 cm rim of surrounding liver parenchyma. Pericystectomy is an ideal procedure for multivesicular cysts and calcified cysts, especially if there is bile duct communication, but is contraindicated for cysts impinging on major hepatic veins or inferior vena.

Some 25–30% of patients with hydatid disease of the liver present with complications, most commonly infection and rupture of the cyst into the biliary tree, although other complications may occur including intraperitoneal or intrathoracic rupture and rupture into hollow viscera and rarely into the vascular tree. The most common infection of hydatid cyst in the liver is by *E. coli* and these infected cysts invariably have bile-stained contents owing to communication with the biliary tree. The clinical features of intrabiliary rupture include jaundice and cholangitis. The diagnosis of intrabiliary rupture is best confirmed by endoscopic retrograde cholangiopancreatography or magnetic resonance cholangiopancreatography. The operative treatment consists of emptying the mother cyst with removal of hydatid membranes and supraduodenal exploration and clearance of the CBD. The biliary communications with the cyst are identified and sutured, following which the choledochostomy is closed over a T-tube. In patients with gross dilatation of the CBD or incomplete clearance or a large biliary communication which cannot be closed, a choledochoduodenostomy is performed. Omentoplasty is an effective alternative treatment if the biliary fistula is small. Omentoplasty has the added benefit of promoting reabsorption of serosal fluid. Internal drainage is a safe and useful procedure and is advised in the presence of a major biliary communication that is difficult to close or when closure of the communication may compromise biliary drainage or when the CBD is grossly dilated. Cystojejunostomy has the dual advantage of internal draining of the biliary fistula and the residual cyst cavity provided the cystojejunostomy is fashioned in a dependent position to avoid stagnation of the cyst contents.

Intraperitoneal rupture is most common with cysts bulging through the anteroinferior surface of the liver; when this occurs, it is accompanied by shock with signs of diffuse peritonitis. These patients require emergency resuscitation followed by surgical treatment with peritoneal lavage using protoscolicidal agents and hypertonic saline. Rarely, intraperitoneal rupture is clinically silent and the patient presents years later with disseminated abdominal hydatidosis.

Amoebiasis

The disease is caused by *E. histolytica*. The parasite lives in the large intestine, causing ulceration of the mucosa with consequent diarrhoea. Secondary lesions may occur, most commonly in the liver, but other tissues may be affected, e.g. lung, brain, genital organs and skin. Amoebiasis has a worldwide distribution but clinical disease occurs most frequently in tropical and subtropical latitudes. It is found in countries where standards of personal and environmental sanitation are low.

Parasitology

The amoeba, which multiplies by binary fission, lives in the lumen of the large intestine where, under suitable conditions, it invades the mucosa/submucosa and ingests red blood cells. When diarrhoea occurs, amoebae are expelled and can be detected in freshly passed fluid stools. In the absence of diarrhoea and when other conditions are favourable for encystation, the amoebae cease feeding, become spherical and secrete a cyst wall, and the nucleus divides twice to form the characteristic mature four-nucleate cyst. This is the infective form and, when ingested, the cyst hatches in the lower part of the small intestine/proximal colon and a four-nucleate amoeba emerges. After a series of nuclear and cytoplasmic divisions, each multinucleate amoeba gives rise to eight uninucleate parasites that establish themselves and multiply in the large intestine.

In endemic areas, the prevalence is stable and the morbidity rate from the disease is low. The disease is spread by cyst passers, who fall into two groups: (1) convalescents who have recovered from an acute attack and (2) individuals who can recall no clinical symptoms or signs of infection. Poor sanitation is more important than climate in the predominance of overt infection in the tropics. Carrier rates of *E. histolytica* among symptomless subjects vary from 20% to 80% in different communities. The parasite can be transmitted by direct contact through contaminated hands of cyst carriers. It is also transmitted indirectly through contaminated food, e.g. raw vegetables.

Pathology

The primary sites of infection in order of frequency are the caecum, colonic flexures, descending colon and rectum. The appendix is sometimes involved and rarely the ileum. Macroscopically, the large intestine becomes studded with discrete ulcers with overhanging edges, the intervening mucosa being relatively normal. The ulcers spread laterally in the submucosa and become confluent. Large areas of mucosa thus become devitalized and form greenish shaggy sloughs. The necrotic process may involve the muscularis propria and even extend to the serosa in severe cases. In patients in whom the host–parasite balance has been altered by drugs, concurrent disease or pregnancy, the entire mucosa may be sloughed. The wall of the bowel is thickened and friable (Figure 11.32).

The amoebae spread laterally beneath the intestinal epithelium and form large 'flask-shaped' or 'water-bottle' ulcers that have overhanging edges and consist of a zone of necrosis surrounded by chronic inflammatory infiltrate of lymphocytes and macrophages with a variable fibroblastic reaction. Amoebae are found at the periphery of the lesion in the submucosa and muscle layers and in the necrotic tissue itself (Figure 11.33).

Liver abscess (Figure 11.34) is the most common extraintestinal complication. The abscess is usually solitary but multiple abscesses are not uncommon. The right lobe of the liver is most frequently affected. Bile appears to destroy the amoebae and for this reason the gallbladder is never affected. In over 50% of patients with amoebic liver abscess, there is no evidence of amoebic infection on stool examination. The amoebae cause lysis of the liver parenchyma, primarily in the periportal region. An expanding area of necrosis ensues and the abscess cavity may reach large proportions. The cavity contains sterile, chocolate-coloured fluid, granular debris and a few inflammatory cells. Amoebae may or may not be present in the pus. Histologically, the wall of the abscess consists of necrotic tissue and compressed liver parenchyma containing a variable infiltrate of monocytes, plasma cells, lymphocytes and fibroblasts.

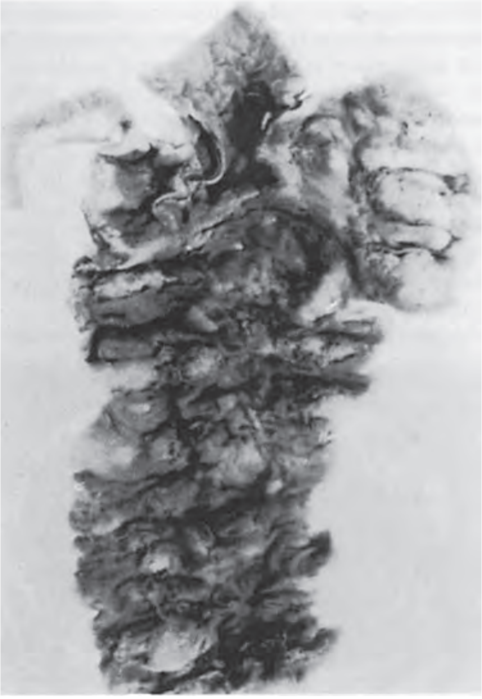


Figure 11.32 Fulminating amoebic colitis resulting from the use of steroids given for a mistaken diagnosis of ulcerative colitis.

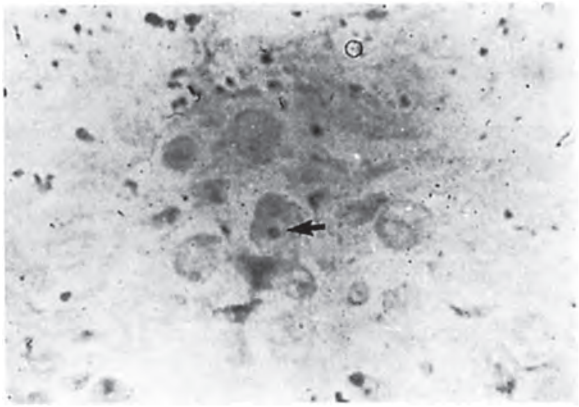


Figure 11.33 Amoebic colitis section of colon. Note *Entamoeba histolytica* trophozoites.

Clinical features

The classic disease is amoebic dysentery, the symptoms of which appear within 1–2 weeks of infection but may be delayed for months to years. The onset is usually gradual, with some looseness for a few days followed by evacuation of 6–12 mucoid, blood-stained motions per day. Colic and tenesmus are unusual unless there is a lesion immediately inside the anus. Physical examination may be negative. Occasionally, during an acute attack, there is palpable thickening of the caecum or left colon with tenderness on palpation. Pyrexia is absent and there is little prostration. The duration of an attack of average severity may be a few days but it may linger on for some weeks. The attack usually subsides spontaneously. There follows a period of remission varying from a few days to several months, even years. During remission, the patient is often constipated. Another attack of dysentery then follows. This sequence of relapses and remissions may continue for several years and is typical of the disease. Complications, e.g. liver abscess, may develop at any time. Complications are encountered in 20% of neglected cases.

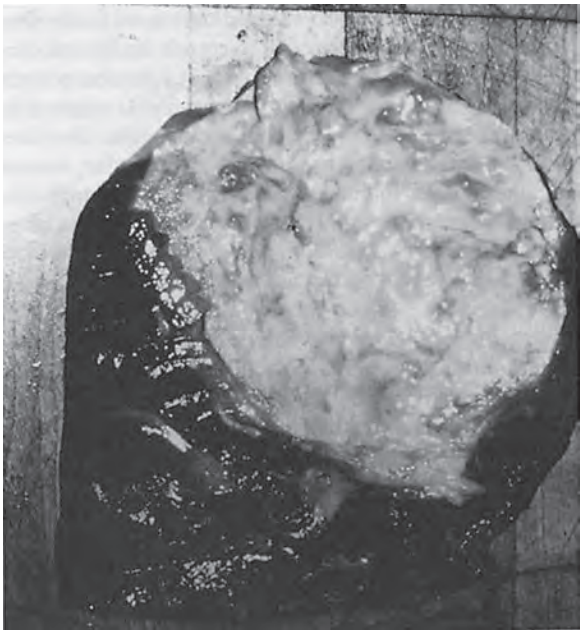


Figure 11.34 Large single amoebic abscess of the liver.

In malnourished individuals or in patients with comorbid disorders, the attacks may be prolonged, severe and even fatal. Fulminating amoebiasis is also encountered in immunosuppressed patients. The fulminating disease has a sudden onset with swinging pyrexia, chills, sweating and very severe dysentery with rapid dehydration and prostration. In such cases the stools are watery with flecks of faecal matter and variable amounts of blood and mucus. There may be severe intestinal haemorrhage or perforation with amoebic peritonitis. The mortality of these cases is high.

The direct complications of an intestinal infection are haemorrhage from erosion of a large intestinal blood vessel, transmural extension of the infection with the formation of amoebic granulomas (amoebomas) and frank intestinal perforation. In addition, a form of slow intestinal leakage through extensively diseased bowel may result in localized peritonitis. This should be suspected in patients whose condition deteriorates and in whom there is increased abdominal distension with signs of ileus. Plain radiology reveals free gas under the diaphragm. Other local complications include amoebic strictures that may occur in any part of the colon and intussusception, which is rare. All of these complications are unusual in the common attacks of average severity.

The clinical features of amoeboma include a palpable mass, usually in the right iliac fossa, low-grade fever, tenderness, and symptoms and signs of intestinal obstruction. Clinically, such a lesion may be indistinguishable from a neoplasm, ileocaecal tuberculosis, appendix mass or actinomycosis.

In the early stages of an amoebic liver abscess, the patient complains of discomfort and fullness in the liver region. The liver enlarges and becomes tender. Moderate fever develops, at first intermittent and subsequently remittent. Sweating is severe, especially at night. The patient becomes anorexic and loses weight. The liver tenderness is maximal over the site of the abscess, usually in the right intercostal region laterally. The enlarged liver may cause obvious bulging of the chest wall and upper abdomen. Chest wall movement is restricted on the affected side. Breathing is painful and the patient develops shallow tachypnoea. The liver dullness is increased upwards and radiology reveals a raised, immobile hemidiaphragm. The liver edge is usually palpable well below the costal margin, and is firm and tender. Jaundice is uncommon. Most patients have a moderate polymorphonuclear leucocytosis. The ESR is raised. Signs of pulmonary involvement (collapse and pleural effusion) are usually limited to the base of the right lung. When the abscess ruptures through the diaphragm into the lung, the abscess contents discharge into the bronchial tree and the patient develops a cough with expectoration of the classic 'anchovy sauce' sputum. Untreated, the abscess may involve other regions, e.g. pericardial sac, peritoneal cavity and parieties. Embolic spread may result in abscess formation in other organs, including the brain. Although usually secondary to liver disease, primary amoebic abscesses of the lung and brain are well documented.

The diagnosis of amoebiasis is established by the identification of *E. histolytica*. Microscopic examination of stool specimens during an acute attack will reveal motile amoebae. In asymptomatic infections and during remissions, the stools contain *E. histolytica* cysts. Repeated stool examinations (minimum of six) should be made before absence of infection can be assumed. Culture methods may assist diagnosis in scanty infections. Sigmoidoscopy often yields useful information. The appearances consist of small yellow ulcers with surrounding hyperaemia, and a normal mucosa in between the ulcers. In chronic cases, amoebic lesions may appear as 'pin-point craters' that are irregularly disposed.

The diagnosis of extraintestinal amoebiasis can be difficult as concomitant dysentery is present in only 5–10% of patients. Posteroanterior chest radiography shows abnormalities in 75% of cases: elevated diaphragm, basal collapse, patchy opacities and pleural effusion. Ultrasound or CT examination of the liver is necessary for establishing the diagnosis and for percutaneous aspiration of the abscess (Figure 11.35).

Serological tests are useful in suspected cases of amoeboma and in those patients diagnosed as having ulcerative colitis who have been to the tropics. The serodiagnostic tests include direct immunofluorescence, ELISA and an isoenzyme electrophoretic technique that differentiates the invasive from the non-invasive form of *E. histolytica*.

Treatment

Medical treatment

Asymptomatic intestinal amoebiasis

In this condition *E. histolytica* is confined to the bowel lumen. Nonetheless, treatment is recommended because invasive disease may develop and because shedding of *E. histolytica* cysts in the environment promotes spread and is thus a public health concern. Treatment is by *luminal* or *contact* amoebicides. There are several such drugs, but few are completely reliable in eradicating amoebae. Diloxanide furoate for 10 days is the most commonly used contact amoebicide, although 10 day courses of Clefamide (2,2-dichloro-*N*-(2-hydroxyethyl)-*N*-[p-(p-nitrophenoxy)benzyl]acetamide) is also used as it is well tolerated and of low toxicity.

Amoebic dysentery

Metronidazole is the standard treatment for acute amoebic dysentery and for amoebic liver abscess. Tinidazole (Fasigyn) is also effective. Metronidazole treatment is usually accompanied by a luminal agent to eradicate colonization. Broad-spectrum antibiotics are added if perforation is suspected.

Amoebic liver abscess

Amoebic liver abscess can be cured without drainage and even by one dose of metronidazole. Clinical defervescence should occur during the first 3–4 days of treatment. Failure of resolution with metronidazole is an indication for surgical intervention.

Disseminated amoebiasis

This is treated with metronidazole (passes the blood–brain barrier).

Surgical treatment

Surgical intervention may be necessary in:

- fulminating amoebic colitis
- amoebic perforation of the colon
- amoebic liver abscess.

Fulminating amoebic colitis

Surgical treatment is indicated when the patient deteriorates despite medical therapy, in the presence of radiological evidence of toxic dilatation, and the onset of an acute episode, e.g. perforation, severe bleeding. The operative treatment consists of resection of the diseased segment with exteriorization of the bowel ends.

Amoebic bowel perforation

This occurs in 1–2% of patients hospitalized for the disease. Although any segment of the small and large intestine may be involved, the most common sites are the caecum, ascending and sigmoid colon. Multiple localized or diffuse perforations are encountered in 25% of cases presenting for surgery. Three types of perforation are recognized: extraperitoneal, perforation of a granuloma or ulcer without acute dysentery, and perforation associated with fulminant disease. Antiamoebic therapy is the primary treatment for all localized extraperitoneal perforations.

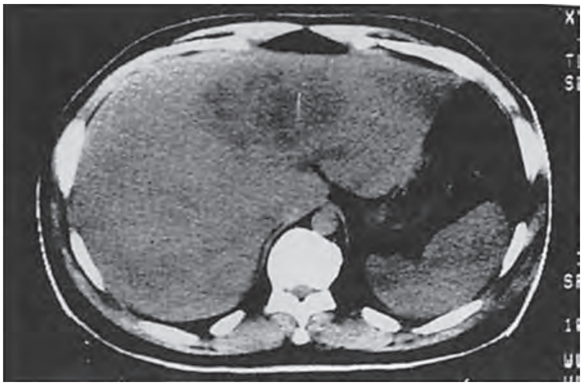


Figure 11.35 CT scan of amoebic abscess of the liver.

Surgery is indicated only if the patient's condition deteriorates and rupture appears imminent. Diversion of the faecal stream by an ileocolic anastomosis beyond any colonic involvement and local drainage is the recommended treatment, except in perforations associated with gangrene and fulminant colitis.

Amoebic liver abscess

The majority of amoebic liver abscesses respond to therapy with metronidazole with or without percutaneous CT or ultrasound-guided aspiration. Surgical treatment is reserved for when medical therapy and percutaneous aspiration have failed and for certain specific indications that include:

- frank or impending rupture
- onset of complications, e.g. secondary infection or haemorrhage
- abscess in the left lobe because of difficulties in aspiration in this area and the risk of rupture into the pericardial sac.

Surgical drainage

In general, patients requiring surgical drainage have advanced disease and the overall mortality of this group is high (30%). A period of therapy with metronidazole lasting for a minimum of 4 days before surgical intervention has been shown to reduce mortality significantly. Transbronchial rupture is usually well tolerated and is often curative, although rarely it may cause fatal pneumonia or lung abscess. Rupture into the pleural cavity is accompanied by shock, respiratory distress and empyema. This situation requires urgent intervention to drain the abscess and the empyema, and to ensure early re-expansion of the collapsed lung. Intraperitoneal rupture requires adequate peritoneal toilet in addition to drainage of the abscess cavity.

Complications following surgical drainage

Apart from respiratory complications and shock, these include liver failure, biliary peritonitis and fistulas. Massive haemorrhage is rare and amoebiasis of the skin is unusual if adequate antiamoebic therapy is started before surgery.

Schistosomiasis

The three species infecting humans are *Schistosoma haematobium, Schistosoma mansoni* and *Schistosoma japonicum*. About 200 million people are affected in various parts of the world. *Schistosoma haematobium* occurs in many parts of Africa, parts of the Middle East and a few foci in southern Europe (Portugal). *Schistosoma mansoni* is found in the Nile delta, Africa, South America and the Caribbean. *Schistosoma japonicum* occurs in China, Japan, the Philippines and other foci in the Far East. Other schistosomes that infect humans include *Schistosoma bovis, Schistosoma matthei* and *Schistosoma intercalatum*.

Parasitology

These worms are trematodes with the peculiar morphology whereby the male is folded to form a gynaecophoric canal in which the female is carried. The adult worms are found in the veins: *S. haematobium* predominantly in the vesical plexus, *S. mansoni* in the inferior mesenteric vein and *S. japonicum* most commonly in the superior mesenteric vein. The female lays eggs, which pass through the bladder or bowel into urine or faeces. A proportion of the eggs remain in the tissues and some are carried to the liver, lungs and other organs. If an excreted egg lands on water, it hatches and produces a free-living form, the miracidium, which swims about by ciliary activity. It next invades an intermediate host, a snail of the appropriate species. Within the snail, it undergoes asexual multiplication, passing through intermediate stages of redia and sporocyst to become the mature cercaria. This is the infective larval stage for humans. It emerges from the snail and swims, being propelled by its forked tail. On contact with humans, the cercaria penetrates the skin, sheds its tail and becomes a schistosomule that migrates to the usual site for mature adults of the species. The respective life cycles are shown in Figures 11.36–11.38.

Humans are the reservoir of *S. haematobium* but naturally acquired infection with *S. mansoni* has been found in various animals, including primates. *Schistosoma japonicum* is widely distributed in various animals (cats, dogs, cattle, pigs, rats, etc.) and these constitute a significant part of the reservoir. Humans acquire the infection by wading, swimming, bathing or washing clothes and utensils in polluted waters. The age and sex distribution of schistosomiasis varies from area to area. One common pattern is of high prevalence rates of active infection in children, who excrete large quantities of eggs, and a lower prevalence of active infection among adults, with the latter exhibiting late chronic manifestations of the disease. The load of infection is an important factor in determining the severity of the pathological lesions and the disease.

Pathology

Schistosoma haematobium

This affects mainly the urinary tract. The bladder lesions include acute and polypoid lesions, fibrous plaques, 'ground-glass' lesions, sandy patches, ulceration, stricture, leucoplakia and cystitis glandularis, fibrosis, calcification of the bladder wall and bladder neck obstruction. However, the bladder may appear normal macroscopically even in fairly severe infections, and mucosal biopsies and press preparations are necessary if *S. haematobium* is suspected.

In the acute stage the bladder may only be hyperaemic, with or without petechial haemorrhages. Ova retained in the vesical tissues, most commonly in the subepithelial layer, result in the formation of pseudotubercles, when the bladder becomes studded with small, yellow, seed-like bodies surrounded by a zone of hyperaemia. They are most frequently situated in the trigone, with the base and lateral walls next most commonly affected. Adult schistosomes are often present in the neighbouring vesical veins. Nodules or polyps may be formed by coalescence of these tubercles, hyperplasia of the mucosa, and early fibrosis and hypertrophy of the muscle. These papillomatous or granulomatous lesions are responsible for the filling defects seen on cystography in the early stages (Figure 11.39). They subsequently shrink to form white fibrous plaques as the ova become calcified. The bladder mucosa may eventually present a ground-glass appearance as a result of the mucosal atrophy and submucosal fibrosis. A fibrocalcific type of polyp is also encountered.

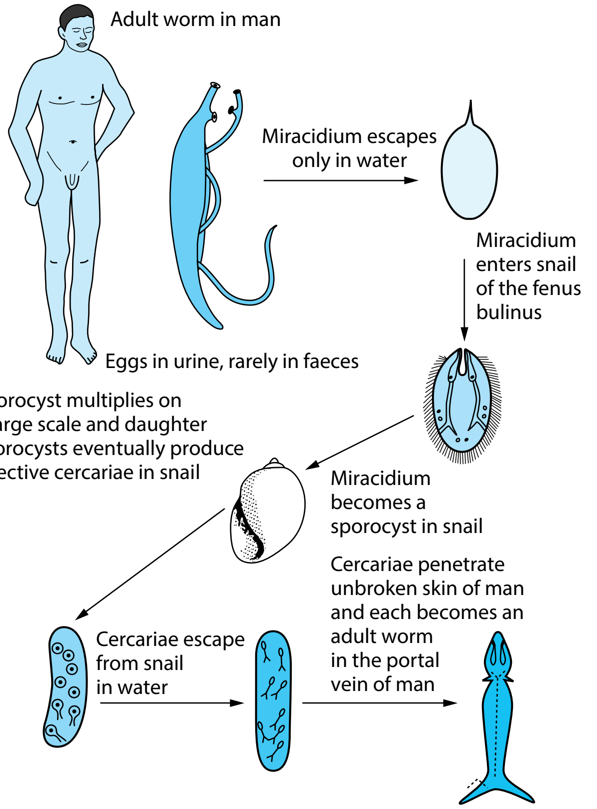

Adult worm in man
Miracidium escapes only in water
Miracidium enters snail of the fenus bulinus
Eggs in urine, rarely in faeces
Sporocyst multiplies on a large scale and daughter sporocysts eventually produce infective cercariae in snail
Miracidium becomes a sporocyst in snail
Cercariae penetrate unbroken skin of man and each becomes an adult worm in the portal vein of man
Cercariae escape from snail in water

Figure 11.36 Life cycle of *Schistosoma haematobium.*

It forms a small, usually solitary lesion consisting of dense fibrous tissue surrounding dilated capillaries and calcified ova and without a covering epithelium. Villous polyps may also form but are less common. They have club-shaped fronds covered by hyperplastic epithelium.

The most common lesion encountered in vesical schistosomiasis is the 'sandy patch'. This late lesion is seen in the trigone, where the mucosa appears roughened, raised and greyish-brown in colour (Figure 11.40). The overlying epithelium may be irregularly thickened or atrophic with areas of metaplasia. In the submucosa and muscularis, pseudotubercles and foreign body granulomas surround ova in various stages of disintegration or calcification, but the predominant feature is dense fibrosis.

The bladder epithelium may undergo atrophy or become hyperplastic. Foci of leucoplakia may be present. Epithelial down growths into the submucosa (Brunn's nests) are a common feature and they may become vesicular (cystitis glandularis). Squamous metaplasia may supervene. Intractable ulceration leads to bacterial infection that may spread to involve the entire urinary tract with abscess formation and septicaemia. The urethra is often involved with stricture formation.

Varying degrees of ureteric involvement, usually the lower two-thirds, are encountered in 70% of cases and are more common in males. Linear calcification in the lower end of the ureter is pathognomonic. In addition, there may be secondary changes in the ureters induced by the bladder lesions, e.g. vesicoureteric reflux and hydronephrosis. Obstructive renal failure may result from ureteric involvement. Renal lesions directly due to the parasite are rare.

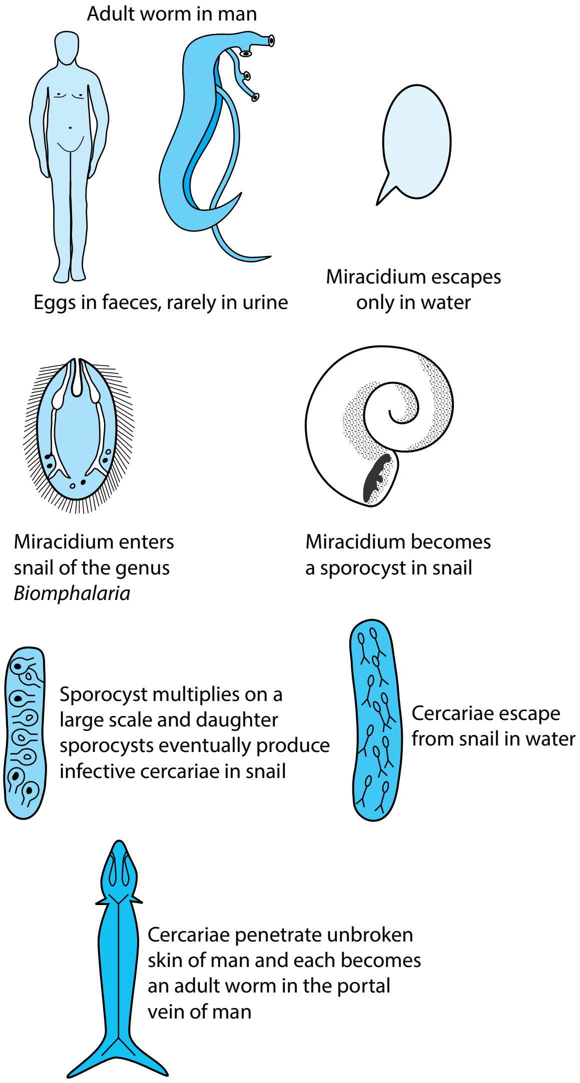

Adult worm in man
Eggs in faeces, rarely in urine
Miracidium escapes only in water
Miracidium enters snail of the genus *Biomphalaria*
Miracidium becomes a sporocyst in snail
Sporocyst multiplies on a large scale and daughter sporocysts eventually produce infective cercariae in snail
Cercariae escape from snail in water
Cercariae penetrate unbroken skin of man and each becomes an adult worm in the portal vein of man

Figure 11.37 Life cycle of *Schistosoma mansoni.*

Schistosoma mansoni and Schistosoma japonicum

Ova are discharged in the faeces but some are retained in the tissues and excite an eosinophilic reaction with the formation of multiple abscesses. In the large intestine, the mucosa is reddened and granular, with pinpoint yellowish elevations surrounded by a hyperaemic zone. Shallow ulcers may form. The inflammatory reaction in the submucosa leads to the formation of sessile or pedunculated polyps. These are found in 17–20% of Egyptian patients. The muscular and serosal layers are also involved with pseudotubercles and these may be associated with a focal peritonitis and intestinal adhesions. Inflammatory masses may be produced in the intestinal wall (bilharzioma). With progressive fibrosis, the intestinal wall becomes rigid and the lumen stenosed. The mesentery is thickened and thrombosis of its veins may occur. Granulomatous lesions may form in the mesenteric and retroperitoneal nodes. These form masses that may simulate a neoplasm (pseudotumour); caecocolic intussusception, intestinal obstruction and rectal prolapse may supervene. With secondary infection, ischiorectal and anorectal abscesses and fistulas may form. Pyloric obstruction has been described in *S. japonicum* infections, and lesions may be found in the stomach, peritoneum

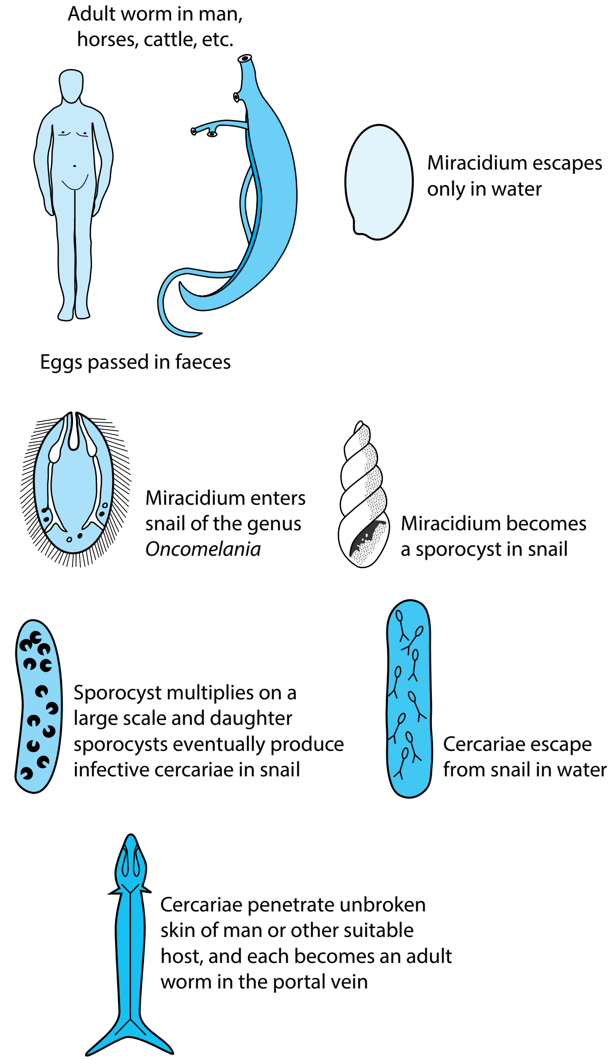

Adult worm in man, horses, cattle, etc.
Miracidium escapes only in water
Eggs passed in faeces
Miracidium enters snail of the genus *Oncomelania*
Miracidium becomes a sporocyst in snail
Sporocyst multiplies on a large scale and daughter sporocysts eventually produce infective cercariae in snail
Cercariae escape from snail in water
Cercariae penetrate unbroken skin of man or other suitable host, and each becomes an adult worm in the portal vein

Figure 11.38 Life cycle of *Schistosoma japonicum*.

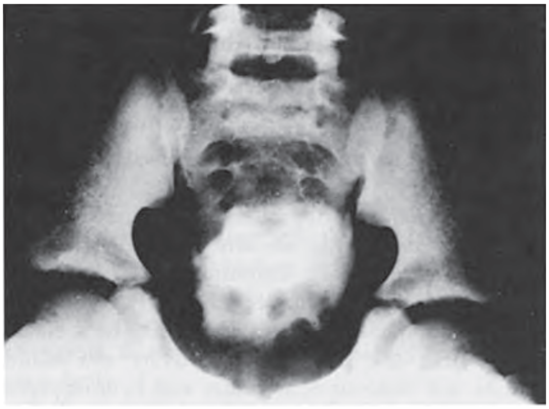


Figure 11.39 Granulomatous lesions in *Schistosoma haematobium* infection seen radiologically as bladder-filling defects.

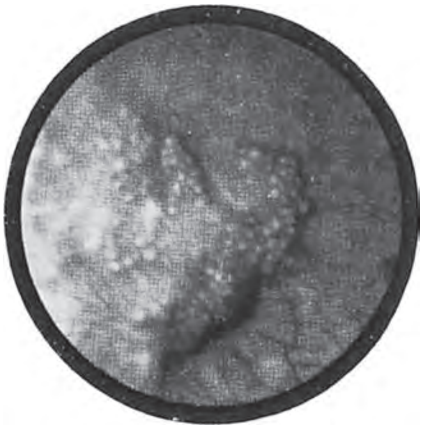


Figure 11.40 Sandy patches of the bladder in *Schistosoma haematobium* infection. (Courtesy of Thieme Verlag.)

and pancreas. Infection of the appendix is common and rarely the small intestine may be more affected than the colon.

As the ova are released into the portal system, the liver is invariably involved. Hepatic involvement is more severe in *S. japonicum* owing to the larger number of ova produced by this schistosome. The liver may be enlarged or contracted and its surface shows a fine to coarse nodularity. Its consistency is firm and the cut surface exhibits white 'clay-pipe stem' fibrosis. This fibrosis is produced by ova in and around the portal venous radicals (Figure 11.41). The ova are usually trapped in the portal venous radicals and only occasionally reach the hepatic sinusoids. The affected portal venules become obstructed from thrombophlebitis. Thin-walled vascular channels form in the fibrosed portal tracts (angiomatoids) and result either from recanalization of thrombosed veins or as adaptive communications between the portal and arterial circulations. An increase in the number and size of the intrahepatic branches of the hepatic artery accounts for the maintenance of a normal liver blood flow in the majority of patients. The parenchymal cells of the liver are usually unaffected and the overall liver function remains good despite established portal hypertension.

The pulmonary changes are usually encountered in the late stages of *S. japonicum* and *S. mansoni* infections. They include widespread fibrosis with the development of cor pulmonale and chronic respiratory failure. Cerebral involvement is more common in *S. japonicum* and spinal disease in *S. mansoni*. Small tubercles are found in the meninges and in the white and grey matter of the brain and spinal cord.

Clinical features

During the stage of invasion and maturation of the worm, which lasts for 12 weeks, the patient may develop a generalized illness with fever, malaise, exhaustion and sometimes diarrhoea and abdominal discomfort. Urticaria may supervene and eosinophilia is invariably present. This initial illness is known as Katayama syndrome and is thought to result from a temporary state of antigen excess that exists until the host's antibody production is mobilized. It may occasionally be very severe and fatal. This stage occurs in all three infections.

Schistosoma haematobium

The stage of established infection occurs 10–12 weeks after cercarial penetration and is manifested by frank haematuria with egg extrusion. The haematuria is intermittent and often transitory. Episodes of haematuria continue to be experienced by untreated patients, although the attacks become less severe

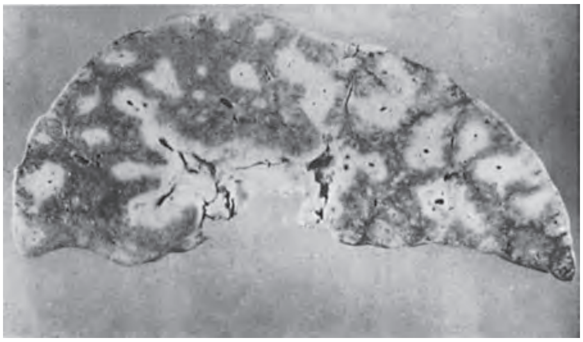


Figure 11.41 Gross appearance of the cut liver surface in *Schistosoma japonicum* infection.

with time. Microscopic examination of urine reveals red blood cells between attacks.

Bladder

Initially, the patient complains of dysuria, frequency and urgency of micturition. Bladder ulceration is accompanied by suprapubic and perineal pain. Two clinicopathological entities have been described: (1) sloughing of polypoid patches in early active disease and (2) chronic ulceration at sites of heavy egg deposition. Increasing frequency accompanies vesical fibrosis and contraction of the bladder. Occasionally, the bladder is large and atonic. This may result from reduced blood supply or bladder neck obstruction.

Kidneys and ureters

Ureteric involvement is common and is usually bilateral. Loin pain results from hydroureter and hydronephrosis. Proximal bacterial infection and stone formation are frequent complications. Pyelonephritis is thus frequent. Two varieties of nephrotic syndrome have been described in Egypt and Brazil. In Egypt, the nephrotic syndrome is related to *Salmonella* septicaemia and the glomerular lesions are proliferative and reversible. *Schistosoma haematobium* infection in Egypt predisposes the patient to the development of a carrier state for *Salmonella typhi* and *Salmonella paratyphi*. In Brazil, the lesions are predominantly membranoproliferative and irreversible with no association with *Salmonella* infection. Bacteriuria is often present in hospital patients and the urine contains eosinophils.

Calculi

These are common in Egypt, where they occur in 25% of infected patients, but are rare in other parts of Africa. They occur in the bladder, ureters, kidneys and urethra and consist of a central core of oxalate surrounded by an outer coat of urate incorporating ova. These stones may disappear following successful medical therapy and this is most likely to happen in young patients in whom fibrosis and irreversible stenosis are absent.

Association with vesical cancer

This is now established. Schistosomal bladder cancer is frequently of the squamous cell type (40–75%) and arises commonly from the trigone and superior apical regions. It is diffuse (multicentric) in one-third of cases. Various theories have been proposed to explain the association but the most favoured is the β-glucuronidase hypothesis. The enzyme β-glucuronidase is excreted in the urine of patients with active *S. haematobium* infection (from the adult worm). The increased urinary β-glucuronidase may hydrolyse inactive carcinogenic glucuronides, releasing active carcinogens. Thus, in Egypt raised levels of urinary metabolites of tryptophan, serotonin and the carcinogen 3-hydroxyanthranilic acid have been demonstrated in the urine of infected patients, and very high levels in patients with established bladder cancer.

Hepatic involvement

The clinical picture is dominated by bleeding oesophageal varices. The spleen is invariably massively enlarged in these patients, who have good liver function despite the portal hypertension (Child A or Child B).

Schistosoma mansoni *and* Schistosoma japonicum

The onset of egg laying in a first infection is accompanied by bloody diarrhoea. The lesions in the colon are variable. There may be segmental roughening of the mucosa with congestion, small ulcers and, in the late stages, sandy patches. Polyps may develop, especially in *S. mansoni* infections. Polyp formation is due to high localized egg burden damaging the muscularis mucosa. These polyps are therefore inflammatory in nature and are thus reversible with medical treatment.

In some cases, severe dysentery with frank ulceration and massive haemorrhage occurs and is often fatal. Perforation and stricture of the colon are uncommon complications. Rectal prolapse is usually associated with polyps. Pseudotumour results from an excessive connective tissue reaction to schistosomal granulomas. It may be located within or outside the intestinal lumen and may be palpable. Pseudotumour often causes stenosis.

Hepatosplenic schistosomiasis occurs in patients with a heavy worm load about 5–15 years after infection. Ova are deposited in the terminal radicles of the portal vein with granuloma formation and fibrosis of the portal tracts leading to portal hypertension, massive splenomegaly and ascites. These patients often have associated HBV infection. *Schistosoma mansoni* infections may be associated with chronic *Salmonella* infection and the nephrotic syndrome.

Diagnosis

Diagnosis is established by the demonstration of living schistosome eggs in the urine (*S. haematobium*) or stools (*S. mansoni, S. japonicum*) or biopsy material. Dead eggs only signify past infection.

Medical treatment

The only broad-spectrum antischistosomal agent that is effective against all *Schistosoma* species is praziquantel. Oxamniquine is only effective against *S. mansoni*, while metrifonate is only effective against *S. haematobium*.

Praziquantel is the drug of choice and is given in the following dosage:

- *S. haematobium*: 30 mg/kg (single dose)
- *S. mansoni* with acute intestinal infection: 20 mg/kg daily for 3 days
- *S. mansoni* with colonic polyposis: 20 mg/kg twice daily for 3 days

- *S. mansoni* with hepatosplenic disease: 30 mg/kg daily for 6 days
- *S. japonicum*: 20 mg/kg three times for 1 day.

Surgical aspects

The need for surgical intervention in schistosomiasis may arise from the development of complications of the disease.

Portal hypertension

Schistosomiasis causes a parenchymal block, usually without significant hepatocellular dysfunction or reduction of the hepatic blood flow. Minor degrees of portal hypertension and hypersplenism may be reversed with anthelmintic therapy. Treatment for oesophageal varices is only required if bleeding has occurred. Many of these patients may be managed with sclerotherapy or banding of the varices. Surgical intervention is indicated if bleeding cannot be controlled in this way, or recurs after repeated sclerotherapy. Portacaval shunts are contraindicated in these patients because of the high risk of severe encephalopathy. Selective decompression by a Warren shunt gives good results. The favoured procedures are, however, either splenectomy with porta-azygos disconnection or the more extensive devascularization operation of Sigiura.

Bilharzial granuloma of the gastrointestinal tract

This complication predominantly affects the large bowel, with small intestinal and appendiceal lesions being uncommon. Small bowel granuloma usually presents with acute intestinal obstruction, or more rarely with mesenteric infarction.

The clinical features of large bowel bilharzial granuloma are varied. The patient may have chronic symptoms, e.g. vague abdominal pain, palpable mass in the lower abdomen, passage of blood and mucus, anaemia and rectal prolapse. In these cases, the diagnosis is usually made by sigmoidoscopy and biopsy. Less commonly, the patients present with large bowel obstruction and at operation a lesion in the sigmoid or descending colon or rectum is found (pseudotumour) that is macroscopically indistinguishable from carcinoma. Resection without primary anastomosis is performed in these patients. In the elective situation after adequate bowel preparation resection with primary anastomosis is safe.

Obstructive uropathy

Cystoscopy should be avoided when possible because of the risk of introducing secondary infection. It is reserved for diagnostic problem cases, e.g. exclusion of neoplasm. The surgical treatment of obstructed ureters remains controversial. An accurate assessment of the extent of the disease and the state of the bladder by intravenous pyelography and ultrasound is essential. Adequate excision of the diseased segment with direct reimplantation or by means of ileal conduits with vesicoileal anastomosis gives the best results. More severe ureteric and bladder disease requires major surgical treatment (ileocaeco-urethroplasty, etc.) in specialized urological units.

Malaria

Malaria is still the most widely spread communicable disease in the topics. About 1 billion of the world's population are still at risk. The two most common parasites are *Plasmodium vivax* and *Plasmodium falciparum*.

Parasitology

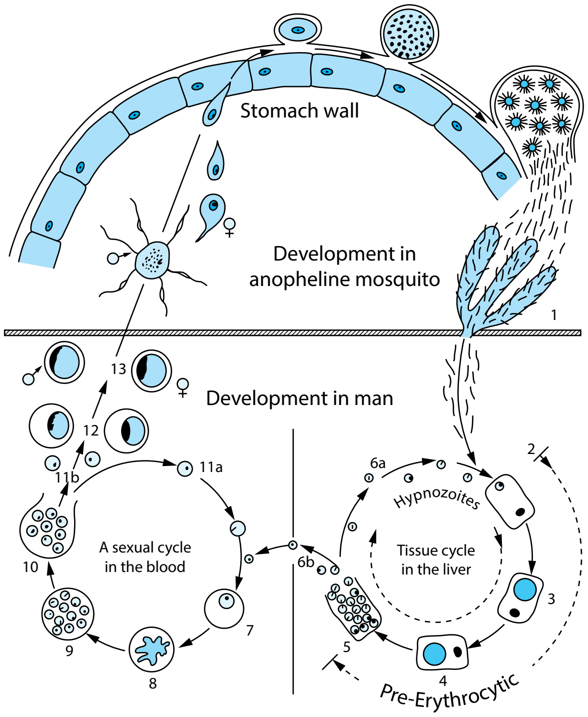
The complete life cycle of the human malaria parasite embraces a period of development within the mosquito and a period of infection in humans (Figure 11.42). After ingestion of infected human blood, a period of development (10–14 days) occurs in the mosquito, resulting in the production of sporozoites. A mosquito bite infects the human host with these forms, which remain in the circulating blood for 30 minutes, then enter tissue cells, notably the liver. During the next 7–9 days, the sporozoites develop in the hepatocytes. This stage of development is known as the pre-erythrocytic cycle. The cryptozoic schizonts that form rupture and release numerous merozoites, most of which enter the circulation to invade the erythrocytes, thus starting the erythrocytic cycle. No symptoms of malaria are experienced during the pre-erythrocytic cycle. The plasmodium first appears in red cells as a small speck of chromatin surrounded by scanty cytoplasm and soon becomes a ring-shaped trophozoite. As the parasite develops, pigment particles appear in the cytoplasm and the chromatin becomes more prominent. Chromatin division then proceeds and, when complete, the mature schizont containing daughter merozoites has formed. The parasitized red blood cells now rupture, releasing merozoites, the majority of which re-enter erythrocytes to reinitiate erythrocytic schizogony. In *P. falciparum* infection, the erythrocytic cycle takes 36–48 hours (subtertian); in *P. vivax* and *Plasmodium ovale* infection 48 hours (tertian); and in *Plasmodium malariae* 74 hours (quartan). The extent of the red cell infection differs from parasite to parasite, being highest (15%) with *P. falciparum*.

In response to an unknown stimulus, a number of the merozoites released when the red cells rupture develop into male and female gametocytes that are inert in the human but provide the reservoir of infection, enabling mosquitoes to perpetuate the malaria cycle. A certain proportion of the merozoites liberated in the pre-erythrocyte phase do not enter the bloodstream. Instead, they re-enter the liver cells to produce metacryptozoic schizonts that are responsible for the persistence of the exoerythrocytic cycle (EE) that occurs in *P. vivax, P. ovale* and *P. malariae* but not in *P. falciparum*. The EE cycle is responsible for the reappearance of malaria some years after clinical cure. The exception to this is *P. falciparum,* in which relapses are not encountered after adequate medical treatment.

Clinical features

Plasmodium falciparum

This is the only type of malaria that may be directly fatal. The clinical picture is extremely variable. A common misconception concerns the periodicity of the fever which, especially in first attacks, is irregular and occurs daily. Headaches, malaise, nausea, vomiting and generalized joint pains are frequent, presenting symptoms of an uncomplicated attack. On physical examination, there is hepatosplenomegaly and a variable degree of anaemia. This rather non-dramatic clinical picture can deteriorate suddenly into one with severe manifestations and a fatal outcome. Abdominal presentation resembling acute appendicitis


Stomach wall
Development in anopheline mosquito
♀
♂
1
13
♀
Development in man
12
11b
11a
2
6a
Hypnozoites
A sexual cycle in the blood
10
6b
Tissue cycle in the liver
3
9
7
5
8
4
Pre-Erythrocytic

This cycle may be repeated many times over a period of months or years

The exo-erythrocytic cycle may be repeated many times over a period of years

Figure 11.42 Life cycle of *Plasmodium vivax* and *Plasmodium malariae.* The life cycle of *Plasmodium falciparum* is similar but hypnozoites do not occur. (1) Sporozoites entering salivary glands of mosquito; (2–5) pre-erythrocytic cycle in the liver; (6a) hypnozoites in liver responsible for initial invasion of erythrocytes and for subsequent relapses; (6b–11a) erythrocytic schizony; (11b) immature gametocytes awaiting ingestion by anopheline mosquito.

can occur. The severe manifestations include cerebral malaria, algid malaria, severe malarial anaemia, hyperpyrexia, jaundice, pulmonary oedema and malarial haemoglobinuria. Patients are usually dehydrated and may have herpes labialis.

Plasmodium malariae

This produces the nephrotic syndrome in children with longstanding infection.

Plasmodium vivax

Fever is the most constant sign. Initially remittent, the pattern changes to intermittent, regularly recurring fever every second day. In children convulsions occur when the fever is high. The classic features of the attack, i.e. cold stage, hot stage and sweating, are unusual in infants and children. General symptoms are similar to those of *P. falciparum* but milder. The spleen enlarges early in the disease, some degree of anaemia may be present and there is often mild leucopenia. The disease is rarely fatal unless the spleen ruptures. A conclusive diagnosis of malaria is made by detection of the parasite in the blood.

Treatment

Treatment consists of antiparasite chemotherapy and supportive management of the various manifestations. Only *P. falciparum* has become resistant to chloroquine. The drugs used in chloroquine-resistant areas include quinine, Fansidar, Metakelfin, mefloquine (Lariam) and halofantrine (Halfan). In areas with resistance to both chloroquine and Fansidar, treatment is with quinine + tetracycline or mefloquine or halofantrine.

Primary chemoprophylaxis is important for visitors to endemic regions and is outlined in Table 11.5. Estimates indicate that every year up to 30 000 travellers contract malaria. Fever developing in a traveller within 3 months of leaving a malaria-endemic area should be investigated urgently. All visitors to endemic malarious areas require prophylaxis. Pregnant women, infants and young children and people who have undergone splenectomy should avoid travel to malarious areas as they are at higher risk. The choice of chemoprophylaxis varies depending on the species and drug resistance prevalent in a country. No chemoprophylaxis regimen provides 100% protection. Thus, an essential component of prophylaxis is the prevention of mosquito bites. Malaria develops up to a year after visiting an endemic area, although it usually develops within 3 months of the visit. The prophylactic treatment should be started 1 week before departure and continued for 4 weeks after returning from the endemic area.

Multidrug-resistant malaria

In areas of Thailand near the borders with Cambodia and Myanmar and in Western Cambodia, *P. falciparum* infection is resistant to chloroquine or pyrimethamine–sulphadoxine, and sensitivity to quinine is reduced. In these areas, chemoprophylaxis with doxycycline is recommended along with rigorous personal protection measures. Doxycycline is contraindicated in pregnant women and children below the age of 8 years.

Filariasis

The more common of the several species of filarial parasites transmitted by various arthropod vectors include *Wuchereria bancrofti, Onchocerca volvulus, Brugia malayi* and *Loa loa.* Some of these essentially tropical and subtropical diseases produce filarial elephantiasis of the lower limbs and genitalia as a result of the occlusion by sclerosis of the lymphatics by adult worms resulting in gross lymphoedema. Many manifestations remain subclinical.

The life history of the various parasites that infect humans follows a common pattern, although details vary from species to species. The adult male and female worms of the infected person reproduce in the skin or lymphatics. The long, thin larvae are known as microfilariae. These enter the peripheral bloodstream at specific hours depending on the filarial species and the feeding habits of the associated arthropod vector. In the insect (mosquitoes, black flies, midges, etc.), the microfilariae mature and migrate to the biting structures of the arthropod and thus enter the bloodstream or skin of the next victim of the insect bite. Further maturation occurs in the human, where the adult

Table 11.5 Primary chemoprophylaxis of malaria

Chloroquine-sensitive areas	Chloroquine 300 mg once weekly, or proguanil 200 mg daily
Chloroquine-resistant areas	Chloroquine 300 mg once weekly and proguanil 200 mg daily, or mefloquine 250 mg once weekly

worms congregate into masses causing granulomatous painful swellings. *Wuchereria bancrofti* is the main cause of elephantiasis and hydrocele.

Pathology

Filarial infection generates a pronounced significant inflammatory immune Th2 response that leads eventually to symptomatic lymphatic obstruction. Symptoms usually do not manifest until adolescence or adulthood, when worm burden is usually the highest. The various conditions caused by these parasites are best considered as (1) lymphatic filariasis, (2) onchocerciasis and (3) loiasis.

Lymphatic filariasis

This arises from infection with *Brugia timori*, *B. malayi* or *W. bancrofti*. The clinical course is broadly divided into (1) asymptomatic microfilaraemia, (2) acute phases of adenolymphangitis and (3) chronic irreversible lymphoedema. Acute adenolymphangitis is characterized by sudden onset of febrile painful lymphadenopathy, which usually subsides within 1 week. However, the acute phase may consist of filarial fever only without the associated lymphadenitis.

Hydrocele is the most common manifestation of chronic *W. bancrofti* infection in males in endemic areas, but is rare with *B. malayi* and *B. timori* infection. Chyluria also may be present in chronically infected persons. However, the main feature of the disease is severe lymphoedema, most frequently in the lower extremities, but it may involve the arms, breasts, scrotum, penis and vulva.

The WHO has developed a grading system for the severity of the lymphoedema:

- Grade 1: pitting oedema, reversible with limb elevation.
- Grade 2: non-pitting oedema, irreversible with limb elevation.
- Grade 3: severe swelling with sclerosis and skin changes.

Onchocerciasis

This is also known as leopard skin or river blindness. Initially, the symptoms result from the presence of microfilariae in the skin and include pruritus, subcutaneous lumps and lymphadenitis. Patients with onchocerciasis may develop impaired vision and even blindness. The hallmark clinical triad consists of dermatitis, skin nodules (onchocercomas) and ocular lesions. Skin lesions include oedema, pruritus, erythema, papules, scab-like eruptions, altered pigmentation and lichenification. Skin nodules are common over bony prominences. Eye lesions are caused by an abnormal host immune response to microfilariae and include punctate keratitis, pannus formation, corneal fibrosis, iridocyclitis, glaucoma, choroiditis and optic atrophy.

Loiasis

This is caused by *L. loa* infection. The manifestations are usually confined to subcutaneous swellings on the extremities, localized pain, pruritus and urticaria. The diagnostic feature of loiasis is a large transient area of localized non-erythematous subcutaneous oedema most commonly around the joints. Other rare features include arthritis, breast calcification, meningoencephalopathy, endomyocardial fibrosis, peripheral neuropathy, pleural effusions and retinopathy.

Treatment

Symptomatic lymphatic filariasis is treated with diethylcarbamazine. Onchocerciasis is treated by diethylcarbamazine and suramin with inpatient care recommended to monitor for reactions and complications of therapy. Infestations with *Mansonella perstans*, which is resistant to standard antiparasitic treatment, is with doxycycline, which kills or sterilizes the filarial nematode. Annual mass treatment with albendazole and ivermectin is used to interrupt transmission of *W. bancrofti*.

Surgical treatment

The treatment of lymphoedema is unsatisfactory. Perhaps the most important practical consideration is the prevention of secondary bacterial infections, attention to hygiene and skin care with prompt antibiotic treatment of bacterial infections when they arise. Some have reported benefit with oral administration of the benzopyrene 5,6-benzo-[alpha]-pyrone in reducing both filarial and non-filarial lymphoedemas of the extremities. A large variety of surgical procedures, which include lymphangioplasty, lymphovenous anastomosis and excision (debulking) of fibrotic subcutaneous tissue, have been used but are of limited value. However, large hydroceles and scrotal elephantiasis can be managed effectively with surgical excision.

Elephantiasis of the lower limbs occurs in certain areas of East Africa and Ethiopia, where there is no filariasis. These cases of non-filarial elephantiasis of the lower limbs are the result of an obstructive lymphopathy of the peripheral lymphatics caused by aluminosilicate and silica absorbed from the soil through the skin of the feet.

GUIDE TO FURTHER READING

Anesti E, Brooks P, Majumder S. Images in emergency medicine. Gas gangrene. *Ann Emerg Med* 2007;**50**:14–33.

D'Souza G, Kreimer AR, Viscidi R, *et al.* Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* 2007;**356**:1944–1956.

Ghany MG, Strader DB, Thomas DL, Seeff LB. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology* 2009;**49**:1335–1374.

Hibbs JR, Frickhofen N, Rosenfeld SJ, *et al.* Aplastic anemia and viral hepatitis non-A, non-B, non-C? *JAMA* 1992;**267**:2051–2054.

Hughes SA, Wedemeyer H, Harrison PM. Hepatitis delta virus. *Lancet* 2011;**378**:73–85.

Irish C, Herbert J, Bennett D, *et al.* Database study of antibiotic resistant tuberculosis in the United Kingdom, 1994–96. *BMJ* 1999;**318**:497–498.

Jayaprakash V, Reid M, Hatton E, *et al.* Human papillomavirus types 16 and 18 in epithelial dysplasia of oral cavity and oropharynx: a meta-analysis, 1985–2010. *Oral Oncol* 2011;**47**:1048–1054.

Koutsky LA, Ault KA, Wheeler CM, *et al.* A controlled trial of a human papillomavirus type 16 vaccine. *N Engl J Med* 2002;**347**:1645–1651.

Malik AA, Bari SUL, Amin R, Jan M. Surgical management of complicated hydatid cysts of the liver. *World J Gastrointest Surg* 2010;**27**:78–84.

Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 2006;**118**:3030–3044.

Peters W, Gillis HM. *A Color Atlas of Tropical Medicine and Parasitology*, 3rd edn. London, UK: Wolfe Medical Publications, 1989.

Pories SE, Gamelli RL, Mead PB, *et al.* The epidemiologic features of nosocomial infections in patients with trauma. *Arch Surg* 1991;**126**:97–102.

Schiffman M, Castle PE, Jeronimo J, *et al.* Human papillomavirus and cervical cancer. *Lancet* 2007;**370**:890–907.

CHAPTER 12

Surgical nutrition

SIR ALFRED CUSCHIERI

Introduction

This chapter describes the causes, recognition and clinical consequences of malnutrition. The principles of artificial nutrition support are discussed. The indications and clinical management of the techniques of enteral and parenteral nutrition are outlined, with particular reference to the surgical patient.

Pathology of protein energy malnutrition

The development of disease may be accompanied by a loss of appetite, intestinal malfunction associated with the impaired absorption of nutrients and in some cases metabolic changes leading to tissue wasting. Thus, in many patients there is an inadequate supply of nutrients which, in combination with the influence of inflammatory mediators, leads to weight loss and defective organ function. This process is described by the term *disease-associated malnutrition* (DRM).

The normal feeding–fasting cycle is characterized by the postprandial deposition of glycogen and fat and protein synthesis, modulated by the secretion of insulin. Subsequently, a reduction in the secretion of insulin and an increase in the secretion of glucagon facilitates the mobilization of glucose from glycogen, and proteolysis and lipolysis to provide energy substrates.

During prolonged starvation the supply of glycogen is rapidly exhausted and glucose, an essential fuel for the brain, is formed in the liver from amino acids derived from muscle protein (gluconeogenesis). As the insulin concentrations fall, lipolysis liberates free fatty acids; these compete with glucose for cellular uptake and are used for the generation of energy by organs including muscle and the liver, thereby minimizing the requirement for glucose. Within the liver, fatty acids are converted to ketone bodies; unlike fatty acids, ketones gain access to the central nervous system and supply some energy to the brain, further reducing the requirement for gluconeogenesis. Thus, the wasting of muscle tissue is minimized. The amount of muscle tissue metabolized for the production of glucose is reduced from 75 g on day 3 to 20 g on day 40 of starvation. Nevertheless, tissue wasting, which progresses during the course of starvation, involves the respiratory muscles, the heart and the intestine, in addition to the adipose stores.

In the previously fit subject death from complete starvation occurs after the loss of 30–40% of the initial body weight, usually by 60–70 days, when organ impairment may be irreversible. In the hospital the rate of nutritional depletion is influenced by several factors. Many patients experience partial starvation when nutritional decline is slower, similar to the Minnesota experiment in which previously fit males were given a 1600 kcal diet for 12 weeks. Patients may receive a diet that is inadequate not only in the energy content but also in the range of nutrients. The historic view that the provision of carbohydrate without protein leads to kwashiorkor, through increased insulin concentrations and the suppressed breakdown of muscle protein, thereby depriving the liver of amino acids for albumin synthesis, is no longer tenable. The deficiency of antioxidant micronutrients, in addition to the generation of cytokines, is the more likely cause of this nutritional disorder. Cytokines, small peptides, notably tumour necrosis factor and insulin-like growth factor 1 (IGF-1), which originate from lymphoid and haemopoietic cells in response to tissue damage, promote lipolysis and proteolysis.

Infections, burns and trauma, including surgical operations, lead to accelerated nutritional decline. Although this has been explained by a presumed increase in energy requirements in the stressed patient, and the humoral response to stress, most of the observed changes are attributable to the release of cytokines. Energy requirements in critical illness have previously been

BOX 12.1 Metabolic effects of cytokines

- Anorexia
- Pyrexia
- Release of amino acids from muscle tissue
- Increased glucose transport
- Stimulation of hepatic lipid synthesis
- Promotion of acute phase protein synthesis
- Reduced albumin synthesis
- Increased vascular permeability
- Immune cell availability

overestimated; any increased energy needs are often matched by a reduction in mobility. Cytokines such as interleukin, tumour necrosis factor and interferons, which are principally responsible for the metabolic changes, share multiple activities, some of which are listed in Box 12.1.

Furthermore, in severely ill patients, especially those in intensive care, the effect of the initial pathology may be compounded by the secondary effect of the breakdown in the gut mucosal barrier, with increased permeability to microbial toxins, and enhanced translocation of bacteria, continuing cytokine generation and leading to greater tissue and organ damage. Data to support the concept that such a breach of intestinal barrier function promotes the systemic inflammatory response in this manner have largely been derived from the animal model.

Disease-associated malnutrition syndrome

The prevalence of DRM in hospital admissions in the UK is high with studies indicating that 40% are malnourished to variable extents and, according to the McWhirter and Pennington study, the nutritional state of many of these patients deteriorates further during the course of hospitalization. In a study of 25 general practitioner practices, 10% of patients in primary care were malnourished. Other surveys have shown that many malnourished patients admitted to hospital go undiagnosed because of inadequate screening (European Nutrition for Health Alliance 2008). DRM results in a decrease in food intake that, if sustained, leads to metabolic changes, which alter body composition and depress function, which become manifest with apparent weight loss, reduced subcutaneous fat and muscle wasting. Chronic malnutrition has adverse effects on virtually all body systems:

- *respiration*: atrophy/weakness of respiratory muscle, decreased lung volumes
- *circulation*: degenerative changes in cardiac muscle, reduced cardiac output
- *haemopoiesis*: anaemia, leucopenia, thrombocytopenia
- *gut*: atrophy of the gut mucosa, decreased gastric acid production
- *immune system*: depressed cellular and humoral immunity
- *skeleton*: decreased bone strength/mass
- *skin*: atrophy, local oedema, increased friability, risk of pressure ulcers
- *kidney*: impaired filtration and excretion.

In surgical practice, DRM increases morbidity and mortality following surgery, especially infective complications (skin and surgical site) and delayed wound healing/dehiscence, and retards patient recovery with increases in hospital stay and costs. Hence patients with significant DRM require preoperative nutritional therapy.

The economic importance of detection of DRM and its treatment by supplements is confirmed by the report of the British Association of Parenteral and Enteral Nutrition *Advancing Clinical Nutrition*. This includes a systemic review/meta-analysis of the published surgical studies [seven randomized controlled trials (RCTs) and one cross-over trial) in patients undergoing abdominal and orthopaedic surgery. This analysis showed a net cost saving in favour of supplementation, when calculations were based on bed-day costs (mean value of £166 per patient, and excess bed-day costs £363 per patient). There was also a significant net cost saving in favour of supplementation when calculations were based on average complication costs (£321 per patient). Six RCTs involving abdominal surgery ($n = 418$ patients) that were amenable to a meta-analysis showed a significant cost saving in favour of supplementation.

An International Guideline Committee under the auspices of the American Society for Parenteral and Enteral Nutrition and the European Society for Clinical Nutrition and Metabolism has produced an aetiology-based classification of DRM that incorporates the current understanding of inflammatory response. The Committee proposed the following nomenclature for nutrition diagnosis in adults:

- *Starvation-related malnutrition*: when there is chronic starvation without inflammation; examples include anorexia nervosa.
- *Chronic disease-related malnutrition*: when inflammation is chronic and of mild to moderate degree; examples include organ failure, pancreatic cancer, rheumatoid arthritis or sarcopenic obesity.
- *Acute disease- or injury-related malnutrition*: when inflammation is acute and of severe degree; examples include major infection, burns, trauma or closed head injury.

In surgical practice, the diseases that are commonly associated with protein energy malnutrition include:

- any disorder that impairs the ingestion, digestion or absorption of food
- diseases associated with a prolonged inflammatory response
- diseases which cause an increased energy expenditure.

Some common examples are listed in Table 12.1. There are usually several factors leading to weight loss in each patient. For example, the patient with chronic obstructive pulmonary disease may have difficulty eating because of dyspnoea and the effect of medication such as theophylline, chronic infection due to bronchiectasis, and increase in respiratory workload associated with airways obstruction.

Consequences of malnutrition

The syndrome of malnutrition is associated with a poor outcome in hospital patients. Prolonged hospital stay and increased morbidity and mortality have been documented in affected surgical patients who are at increased risk of infection and poor wound healing.

Table 12.1 Some causes of disease-associated malnutrition

Anorexia	Depression Anorexia nervosa, bulimia Chronic disease
Inability to eat	Neurological disorders Oropharyngeal disease Oesophageal disease
Intestinal disease	Inflammatory bowel disease Gluten enteropathy Radiation enteropathy Hollow visceral myopathy
Catabolic response to trauma	
Specific infections	AIDS Tuberculosis

The adverse consequences of malnutrition stem from (1) protein energy deficit, (2) vitamin deficiencies and (3) deficiencies of trace elements.

Protein energy malnutrition

Malnutrition impairs organ function (Table 12.2). Mental function can be affected, and patients become withdrawn and apathetic. Malnourished patients are often reluctant to co-operate with their treatment. Muscles become weaker and fatigue more rapidly. These effects can be demonstrated in normal people during dietary restriction of only 2 weeks' duration. Not only may such changes affect the mobilization of patients after operative procedures and illness, they may also contribute to respiratory and cardiac failure.

There is a failure of antibody production and phagocyte function. Other aspects of the immune response may be impaired; thus, malnourished patients are at increased risk of infection. Thermoregulation is also defective and this may lead to hypothermia, especially in the elderly.

Table 12.2 Some of the effects of protein energy malnutrition

Impaired mental function	Apathy Fatigue Inability to co-operate with treatment
Impaired muscle function	Respiratory failure Delayed mobilization
Impaired immune function	Increased incidence of infection

Children are particularly vulnerable to the effects of starvation, which results in growth failure. Protein energy malnutrition in children is associated with thymic atrophy and similar changes affect other lymphoid tissues.

Vitamin deficiencies

Some patients demonstrate features of specific nutrient deficiencies, in particular vitamin and trace element deficiencies. Examples of vitamin deficiencies include thiamin and Wernicke's encephalopathy, folate and megaloblastic anaemia, and ascorbic acid and scurvy. Thiamin has an important role in the metabolism of carbohydrates and alcohol. Patients who misuse alcohol and eat a poor diet are at particular risk of deficiency, which might be exposed by the use of a glucose infusion in the fasted hospital patient. Vitamin replacement should precede carbohydrate loading in such patients. Examples of the role of vitamins and the effects of vitamin deficiencies are summarized in Table 12.3.

Deficiencies of trace elements

Deficiencies of trace elements may also lead to specific features, e.g. zinc depletion is associated with skin rashes, impaired sense of taste and immune dysfunction; copper deficiency can cause a microcytic anaemia; and selenium depletion leads to a myopathy, which also affects cardiac muscle (Table 12.4). Cardiomyopathy due to selenium depletion was identified as the cause of premature deaths in the Keshan province in China, where the soil content of selenium is very low. Other deficiencies became apparent when patients first receive prolonged parenteral nutrition without adequate micronutrient supplementation.

Assessment of nutritional status

There is evidence that malnutrition is common in hospital patients and that nutritional status deteriorates during hospital stay. The nutritional status of the patient is not assessed in

Table 12.3 Some examples of the roles of vitamins and the clinical effect of vitamin deficiencies

Vitamin	Biochemical function	Clinical features of deficienc
A	Tissue growth and differentiation	Xerophthalmia, impaired dark adaptation
D	Calcium absorption	Rickets in children, osteomalacia in adults
E	Membrane antioxidant	Haemolytic anaemia, neuropathy, myopathy
K	Synthesis of coagulation factors and osteocalcin	Coagulation defects, possible bone disease
B_1 thiamin	Decarboxylation in carbohydrate, fat and alcohol metabolism	Beri-beri: cardiac and neurological effects, Wernicke–Korsakov syndrome
B_2 riboflavin	Oxidative metabolism	Lesions of the mouth and skin
B_6 pyridoxine	Transamination of amino acids	Anaemia, lesions of the lips and skin
B_{12} cyanocobalamin	Recycling of folate coenzymes	Megaloblastic anaemia, demyelination
C	Antioxidant, iron absorption, collagen synthesis	Impaired wound healing, scurvy
Biotin	Carboxylase reactions in lipogenesis and gluconeogenesis	Scaly dermatitis, hair loss
Folate	Purine and pyrimidine metabolism	Megaloblastic anaemia, growth retardation
Niacin	NAD/NADH in oxidative metabolism	Pellagra: rash, weakness and diarrhoea

Table 12.4 Some examples of the functions of trace elements and the clinical effects of deficiencies

Trace element	Biochemical function	Effect of deficienc
Chromium	Insulin activity, lipoprotein metabolism	Glucose intolerance, weight loss, peripheral neuropathy
Copper	Cytochrome oxidase, superoxide dismutase, encephalins	Hypochromic anaemia, neutropenia, cardiac dysrhythmia
Fluoride	Bone mineralization	Dental caries
Iodine	Thyroxin	Cretinism in children, hypothyroidism in adults
Iron	Haemoglobin–myoglobin–cytochrome system	Anaemia
Manganese	Mitochondrial superoxide dismutase	Lipid abnormalities, anaemia
Selenium	Glutathione peroxidase	Myopathy, pseudoalbinism
Zinc	Enzymes of intermediary metabolism and protein synthesis	Diarrhoea, skin rashes, immune deficiency, growth retardation

routine clinical practice and malnutrition is not recognized in the majority of affected patients. Thus, many patients do not receive appropriate nutritional management. Recognition of nutritional depletion is made difficult by the lack of suitable clinical indices of malnutrition and consequent functional impairment.

Malnutrition is identified after clinical assessment supported by anthropometric measurements and some laboratory investigations. An adequate history includes enquiry about the appetite and changes in the diet and the normal body weight. Examination may reveal evidence of muscle wasting as well as features of micronutrient deficiency, such as bruising, rash and glossitis.

Measurement of body composition

The traditional method of measuring body composition from body density determined by underwater weighing and the use of sophisticated techniques such as dual-energy X-ray absorptiometry are inappropriate or unavailable for routine clinical practice.

Anthropometric measurements

The measurements of weight and height are important. Fluctuation in body weight in the long term may reflect nutritional changes, and in the critically ill patient acute weight change is caused by changes in fluid balance. The measurement of height will determine growth velocity in the child, a sensitive marker of nutrition and disease. In the adult, knowledge of height and weight allows the calculation of the body mass index (BMI) from the [weight (kg)]/[height (m)^2]. The normal range for BMI is between 19 and 25. BMI values below 19 represent malnutrition; patients with a BMI in excess of 25 are overweight; and those with values of 30 are described as obese.

Fasting in the patient who is not metabolically stressed leads to preferential mobilization of fat stores that are measured by skin-fold calipers. In the stressed patient proteolysis leads to muscle wasting measured by calculating the mid-arm muscle circumference (MAMC). The MAMC (cm) is derived from the mid-arm circumference – triceps skin-fold thickness × 0.314.

These measurements are useful for the identification of malnourished patients and monitoring patients who are at nutritional risk as well as those who are receiving long-term nutritional support. Variation in technique between observers and changes in the patient's hydration status can also influence the measurements. Furthermore, in relation to the effect of nutritional depletion and possible need for nutritional intervention, it is the change in nutritional status over time, rather than one single measurement, that is important. For example, a preoperative patient may have suffered from an involuntary weight loss of more than 10%, with an attendant increased risk of operative morbidity, yet still have a BMI within the normal range.

Bioelectrical impedance analysis

Bioelectrical impedance analysis is a non-invasive method of assessment that depends on the difference in electrical conductivity of fat and fat-free mass. The impedance of the body to an electrical current is measured, and is assumed to be proportional to the square of the height of the subject divided by the volume. The resistance between the right wrist and right ankle is measured and used to calculate the conductivity. From the impedance, total body fat and body water are calculated. However, the technique assumes a normal hydration state. This may not be the case in many patients, although there have been attempts to produce prediction equations for different disease states, including general surgical patients. Further validation is needed before this technique can be accepted in routine clinical practice.

Measurement of organ function

Malnutrition leads to impaired muscle strength and increased fatigability. The laboratory methods employed for these measurements using graded electrical stimuli are not suitable for clinical practice. Studies have shown that impaired muscle function can be reversed with nutritional support before improvement in clinical nutritional indices is observed. This observation led to the reappraisal of handgrip dynamometry (Figure 12.1). Handgrip dynamometry, determined by the highest value of three readings recorded with a dynamometer gripped in the non-dominant hand, will give a value that reflects nutritional state when compared with values standardized for age and sex. However, other influences, notably the willingness of the patient to co-operate, reduce the specificity and thus the clinical usefulness of the technique. Tests of respiratory function have also been used for the determination of nutritional status. Unfortunately, the results are influenced by three variables: nutritional state, co-operation and pulmonary disease.

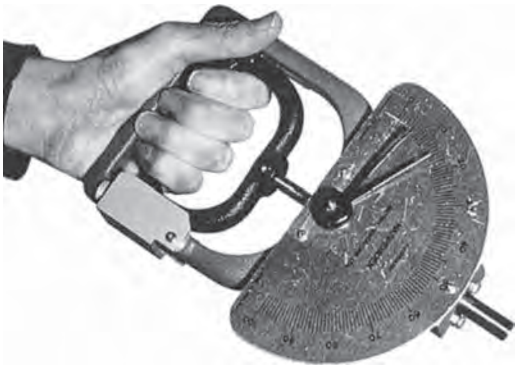


Figure 12.1 Handgrip dynamometer.

Malnutrition leads to anergy, with the loss of cutaneous responses to antigens traditionally determined by the Mantoux response. Many diseases have the same effect. The total lymphocyte count is depressed in malnutrition and has been used to monitor nutritional support. These measurements are little used in this context.

Laboratory investigations

Protein calorie malnutrition causes a decrease in the rate of albumin synthesis. Traditionally, the serum albumin concentration has been advocated as a nutritional marker. Whereas there is evidence to support the use of albumin as a prognostic indicator, in that patients with low serum albumin values have prolonged hospital stay and increased morbidity and mortality, albumin correlates poorly with nutritional status. In the Minnesota experiment the total circulating albumin was reduced by only 2% after 24 weeks of reduced protein and energy intake. This was associated with a 10% reduction in serum albumin concentrations, which may have reflected other influences.

Children with marasmus and adults with anorexia nervosa maintain serum albumin concentrations until the terminal stages of their illness. Conversely, in well-nourished patients who become septic the albumin concentrations fall rapidly, reflecting changes in vascular permeability and fluid retention, as well as a reduction in albumin synthesis caused by the influence of the cytokine responses to infection or tissue damage.

Albumin has a very long half-life of 21 days. Consequently, other proteins have been advocated for diagnosis and nutritional monitoring. They include transferrin, thyroxin-binding prealbumin and retinol-binding protein, with respective half-lives of 8 days, 2 days and 12 hours. These are also influenced by other factors, such as the iron and vitamin A status, and the acute phase response. The IGFs are a family of low-molecular-weight peptides produced by the liver. Studies have suggested that IGF-1, which has a half-life of a few hours, may be a useful marker of nutritional status. Reduced serum concentration may reflect a general decrease in protein synthesis with malnutrition and, unlike the other factors, it is thought not to be influenced by the acute phase response. Currently, this is an expensive assay that requires further validation to define any role in clinical practice.

Clearly, there is no single laboratory marker of nutritional status. However, the laboratory has an important role in the identification of single nutrient deficiencies. The serum values of iron, calcium, magnesium and potassium are commonly measured, along with vitamins such as folic acid, vitamin B_{12} and 25-hydroxyvitamin D, as a guide to replacement during nutritional support, in depleted patients, and in specific disorders of absorption. The measurement of other micronutrients, such as selenium, copper, chromium and manganese, is usually only available in supraregional laboratories. This information is required in severely depleted patients who are receiving prolonged nutritional support, including patients treated by home parenteral nutrition. There is often difficulty in interpretation, blood values do not necessarily correlate with body content and serum values may fall in the context of an acute illness.

Principles of artificial nutrition support

Artificial nutrition support (ANS) may be employed to prevent starvation or to treat malnutrition in patients who are unable to eat or to digest and absorb sufficient nutrients. The potential role of ANS in modifying the inflammatory and immune responses to disease through the provision of specific nutrient substrates – ‘immunonutrition’ – is currently under evaluation.

The need for ANS should be recognized early because impaired organ function can be demonstrated within 2 weeks of starvation, the restoration of nutritional status in the depleted patient is a prolonged process and nutritional repletion may not be achieved in the context of significant illness. ANS may be delivered by the enteral route by oral supplements or enteral tube feeding, or intravenously by parenteral nutrition. Enteral nutrition should be employed wherever possible; it is cheaper, safer and more physiological than parenteral feeding. The intestine has an important immune and barrier function. Luminal nutrition appears to be important for the stimulation of gastrointestinal motility, secretory immunoglobulin (Ig) A and the integrity of the mucosal barrier. Glutamine and short-chain fatty acids are important substrates for the enterocyte and colonocyte, respectively; they are not given in standard parenteral nutrition.

Nutritional requirements depend on the nutritional status of the patient, the nature of the underlying disease and the adequacy of organ function. Any nutrition derived from the oral diet must also be taken into consideration and during enteral feeding the effect of intestinal impairment on energy needs should be considered. In patients who are malnourished some allowance is needed for nutritional repletion, unless the patient is metabolically stressed as a result of sepsis or trauma. The aim in these patients is to maintain nutritional status or minimize tissue loss; repletion is achieved during the recovery phase of the illness. Initial hypocaloric feeding should be considered in patients who are severely wasted to minimize the risk of the refeeding syndrome (see below).

There is no satisfactory bedside method for the measurement of energy requirements. Approximate values can be estimated from the Schofield equation and adjusted for stress, activity and desired energy balance to reflect nutritional status. For the majority of patients 25–35 kcal and 0.2 g nitrogen/kg is sufficient; as a rule, provision should not exceed 40 kcal and 0.3 g nitrogen/kg in the adult. Some patients in the intensive care unit receive lipid energy through medication with propofol, which must be considered when calculating energy needs. The previous views about large energy needs were based on inaccurate methodology; furthermore, it would seem that any increased energy needs associated with illness are balanced by a reduction in energy expenditure through reduced mobility. Excess nutrient delivery leads to futile cycling with lipogenesis, increased production of carbon dioxide, and the consequences of hyperglycaemia and hyperlipidaemia. Cyclical feeding, in which there is an interval free of infusion, has physiological advantages. In comparison with continuous feeding it reduces sodium and water retention, and it may also lead to less fat accretion by avoiding the continuous stimulation of insulin.

Some patients with intestinal disease are able to maintain energy and nitrogen balance, but not electrolyte balance. Patients with Crohn disease frequently develop hypomagnesaemia, and patients with a high jejunostomy become salt and water depleted. This reflects the leaky junctions in the jejunal epithelium, which allow sodium to pass into the intestinal lumen down a concentration gradient. Such patients need additional fluids and electrolytes. Some can be managed satisfactorily with isotonic solutions of sodium and carbohydrate, in which the sodium concentration is approximately 100 mmol/L, provided they avoid drinking hypotonic solutions without sodium, which increase the stomal losses. However, parenteral administration may be required.

Most of the enteral feeds contain the average daily recommended requirement of micronutrients, vitamins and trace elements in a volume of 2 litres. Some commercial trace element and vitamin preparations are compounded with the other nutrients, amino acids, glucose, lipids and electrolytes, for parenteral nutrition. The need for additional supplements in depleted patients should be considered. For example, malnourished alcoholic patients will need additional thiamin, while very malnourished patients who receive nutritional support may require additional phosphate, potassium and magnesium. Electrolyte requirements are also influenced by the administration of drugs such as amphotericin, which can increase the renal excretion of potassium and magnesium, and the presence of impaired organ function, including liver cardiac and renal disease. Under these circumstances the parenteral prescription will need adjustment or special enteral products may be used (see below).

Enteral nutrition

Oral supplements

When used correctly and taken between meals, oral supplements can substantially increase nutrient intake without reducing the consumption of the oral diet. Supplements are useful in anorectic patients and patients with dysphagia pending treatment of oesophageal disease. They have also been shown to reduce morbidity in elderly patients with fracture of the neck of femur, and following abdominal surgery.

Most supplements are milk based, while some are based on soya protein and are fruit flavoured. Not all are nutritionally complete in relation to micronutrients. A complete supplement should be prescribed if it constitutes a substantial proportion of the nutritional intake. The dietician selects the product according to the taste preferences of the patient. Nevertheless, compliance remains a problem, especially in patients with malignant disease.

Enteral tube feeding

Enteral tube feeding is needed in:

- patients with profound anorexia
- patients who are unable to eat or swallow
- some patients with impaired intestinal function.

Anorexia may be a feature of many diseases. Such patients are unable or unwilling to ingest sufficient food or supplements. However, after the initiation of enteral feeding the appetite frequently improves. Oropharyngeal disease, cerebrovascular disease and motor neurone disease are common indications for tube feeding. Tube feeding may also be useful in the postoperative or critically ill patient with gastric paresis, when jejunal feeding may need to be combined with gastric aspiration. Supplemental nocturnal tube feeding is a useful method of exploiting the residual intestinal function in patients with Crohn disease and to increase nutrient intake in patients with cystic fibrosis.

Intestinal access

Access to the intestine will depend on the duration of feeding and gastric function. In the short term and in critically ill patients, fine-bore nasal tubes are convenient. In the long term and for the mobile patient, percutaneous tubes are more suitable. Patients with disease of the oesophagus, impaired gastric function and recent surgery to the upper alimentary tract will need access to the jejunum. The routes of access are summarized in Table 12.5.

There is increasing use of percutaneous endoscopic gastrostomies (PEGs) in patients with chronic neurological disease; once the track has formed the PEG tube can be replaced with a button gastrostomy (Figure 12.2), which is more convenient for the mobile patient. The use of jejunostomy tubes following oesophagogastric surgery can avoid the need for the more expensive and hazardous option of parenteral nutrition.

Nutrient solutions

A wide range of nutrient solutions is available for enteral tube feeding and the types of feed are summarized in Box 12.2.

Table 12.5 Intestinal access for enteral tube feeding

Route of entry	Tip position	Method of placement
Nasal	Nasogastric	Patient or nurse
	Nasojejunal	Endoscopist or surgeon
Percutaneous	Percutaneous gastrostomy	Endoscopist, radiologist or surgeon
	Percutaneous gastrojejunostomy	Endoscopist
	Percutaneous jejunostomy	Surgeon or endoscopist

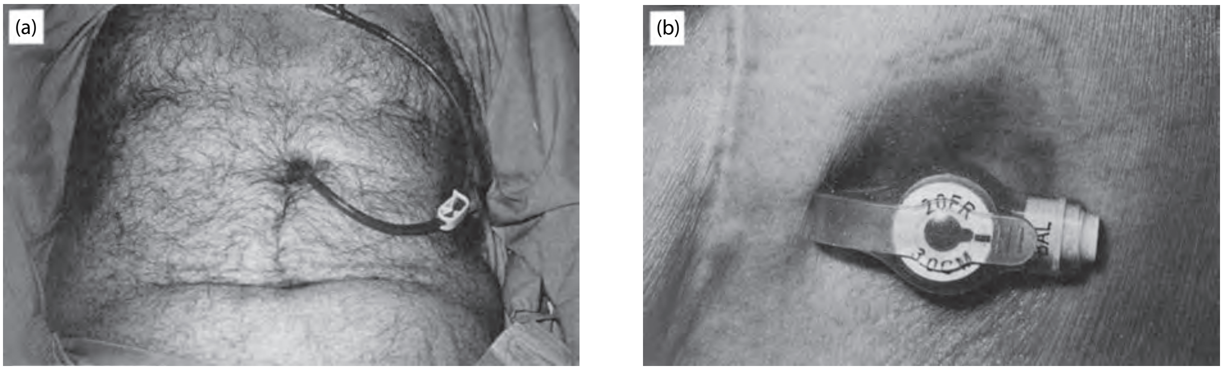


Figure 12.2 (a) Percutaneous endoscopic gastrostomy; (b) button gastrostomy.

BOX 12.2 Types of enteral feed solution

- Oral dietary supplements
- Polymeric feeds
- Predigested chemically defined
- Specialized diets
- Disease-specific feeds

The cheaper whole-protein polymeric diets are preferred for the majority of patients. They comprise whole protein, hydrolysates of starch arid long-chain triglycerides (LCTs). LCTs contain linoleic acid, which is metabolized to arachidonic acid, the precursor of series 2 prostanoids and series 4 leukotrienes that induce inflammation and increase immunosuppression. In contrast, the omega-3 fatty acids lead to the production of series 3 prostanoids and series 5 leukotrienes that have anti-inflammatory and immune-enhancing effects. Some special diets are now available that contain omega-3 fats and additional glutamine. Clinical benefit has been demonstrated in the context of burn injury. Other special diets have a reduced sodium content and are useful in some patients with sodium retention associated with cardiac and liver disease.

Chemically defined diets contain peptides instead of whole protein, and some also contain medium-chain triglycerides. They are used to promote absorption in patients with severe intestinal disease. Disease-specific diets include formulas enriched in branched-chain amino acids for patients with portal systemic encephalopathy due to liver disease, and diets in which a large proportion of the non-protein energy is supplied as lipid for patients with respiratory failure. The clinical evidence to support the use of these disease-specific products is not very convincing.

Nutrient delivery

The need to deliver the nutrient solution at the appropriate rate and to avoid bacterial contamination by the use of clean procedures merits emphasis. Infusion is commenced at 50 mL/h and increased, according to tolerance, to 100 mL/h in the majority of patients who are receiving gastric tube feeding. An initial rate of 25 mL/h may be preferred for critically ill patients, and when initiating feeding in the postoperative patient via a nasojejunal tube. Conversely, flow rates of up to 180 mL/h may be employed in some patients who are receiving enteral tube feeding at home.

Enteral feeding is given by enteral pump infusion, rather than bolus feeding, which is associated with increased gastrointestinal intolerance. When gastric feeding is considered and gastric motility is in question gastric residual volumes are measured. Volumes of up to 400 mL in the critically ill patient do not necessarily preclude gastric feeding, but the volumes should be checked 2 hours after initiating the infusion. This is not necessary with postpyloric feeding. The feeding is given overnight or for longer periods. One further theoretical advantage is that it allows the gastric pH to fall when the buffering effect of the infusion is withdrawn, thus minimizing the tendency to gastric colonization, which might be a factor in the development of infection.

Contamination of the enteral feed should be avoided by the use of a commercial feed, selection of an appropriate reservoir and giving set, and the observance of a protocol when administering the feed. Episodes of infection in the critically ill have been directly related to enteral feeds, and there is the risk of infection with specific intestinal pathogens such as *Salmonella* spp., bacterial translocation in the presence of impaired intestinal and immune function, and bacterial overgrowth, which reduces the efficiency of the feed.

Complications

The complications of enteral feeding can be considered in three groups:

1. nutritional and metabolic
2. complications of nutrient delivery
3. gastrointestinal complications.

Patients require monitoring with respect to fluid and electrolyte status, biochemical parameters and nutritional progress. In particular, fluid retention and electrolyte balance may require attention; hyperglycaemia and hyperkalaemia may be encountered in diabetic subjects and patients with renal impairment. Hypophosphataemia may occur during the refeeding syndrome in the severely malnourished patient. If it is not recognized, the patient may present with thrombocytopenia, cardiac dysrhythmia and mental confusion. Hypomagnesaemia and hypokalaemia are other features of this syndrome, which may reflect rapid cellular uptake with the switch of energy source from endogenous lipid in the starving patient to exogenous carbohydrate. Thiamin deficiency precipitated by glucose infusion represents another example of this problem and is especially relevant in the alcoholic subject.

Complications of delivery are common, especially with nasogastric tubes. More than half of these tubes become displaced and there is a risk of incorrect positioning with the tube in the bronchial tree. Pulmonary aspiration occurs with too rapid infusion or impaired gastric emptying. Complications of the percutaneous tubes include the complications of placement, peritonitis if there is a leak of feed or intestinal contents through the incorrect appositioning of the alimentary tract and the abdominal wall at the site of the tube, and stomal infection.

Diarrhoea is a common problem. Possible causes include bacterial infection of the feed, concomitant antibiotic administration and hypoalbuminaemia. Continuous intragastric infusion of feed has been shown to cause a secretary response in the ascending colon, which is thought to reflect the absence of a normal cephalic phase in tube feeding with an absence of the normal stimulated postprandial peptide YY response. This colonic secretion can be overcome by the instillation of short-chain fats in the caecum; these are normally produced by the action of anaerobic bacteria on dietary fibre. This provides a rationale for the use of fibre-containing feeds and one explanation for the association of antibiotics with diarrhoea. The use of intermittent feeding may allow the gastric pH to fall when the buffering effect of the feed is withdrawn for a period. This may be helpful in reducing gastric and consequently intestinal bacterial colonization, although adequate data to support this suggestion are lacking. Nevertheless, intermittent feeding will minimize fluid and fat accretion in comparison with continuous feeding owing to the effect on insulin secretion.

Parenteral nutrition

The term total parenteral nutrition is frequently used. This implies that the complete range of nutrients is being administered by the parenteral route, which is normally the case, and that nutrition is being delivered exclusively by vein, which is often not the case. Consequently, the term parenteral nutrition is preferred. Parenteral nutrition is often employed to supplement enteral nutrition when there is limited intestinal tolerance or function.

Indications

Examples of common indications for short-term and long-term parenteral nutrition are given in Boxes 12.3 and 12.4.

In the majority of these patients parenteral nutrition will be supplemental. Whereas many patients with wasting due to AIDS have been treated with long-term parenteral nutrition, there is minimal evidence of benefit. Patients with untreatable cancer should only be considered for parenteral nutrition if they are not terminally ill and when the problem is due to untreatable intestinal obstruction and not simply cachexia.

BOX 12.3 Some indications for short-term parenteral nutrition

- Severe inflammatory bowel disease
- Mucositis following chemotherapy
- Severe acute pancreatitis
- Some patients with multiorgan failure
- Following major excisional surgery
- Prior to major surgery in malnourished subjects

BOX 12.4 Some causes of intestinal failure in adults, for which prolonged parenteral nutrition may be needed

- Inflammatory disease
- Crohn's disease
- Gluten enteropathy
- Radiation enteritis
- Motility disorders
- Hollow visceral myopathy
- Scleroderma
- Short bowel syndrome
- Mesenteric infarction
- Miscellaneous disorders
- Inoperable intestinal obstruction
- AIDS

Venous access

The realization that nutrient needs are less than they were previously considered to be, the recognition that most patients receive parenteral nutrition in hospital for less than 2 weeks and the availability of lipid-containing nutrient mixes have led to the use of peripheral veins for the administration of parenteral nutrition. Access may be by conventional Venflon, which should be resited every 1–2 days before thrombophlebitis occurs, or 15 cm ultrafine catheters, which are inserted in one of the antecubital veins (Figure 12.3). The peripheral line should be reserved exclusively for the administration of nutrition and it requires the same standard of care as a central line. Under these circumstances peripheral access will last for 2 weeks or longer.

Central parenteral nutrition is needed when prolonged treatment is envisaged, there are unusual nutritional requirements with volume restriction and a hyperosmolar solution, or in the

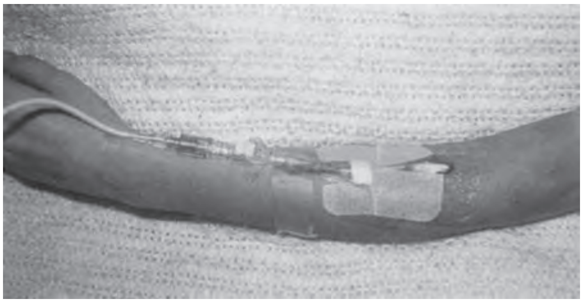


Figure 12.3 Peripheral feeding catheter with Interlink connection device.

absence of adequate peripheral veins. Central access is normally via tunnelled central catheters. For mobile patients and long-term treatment of more than 2 months, cuffed catheters are preferred (Figure 12.4). Some of the latter patients who need home parenteral nutrition may favour an implanted subcutaneous port. More recently, the peripherally inserted central catheter has gained popularity. Central catheters should be placed with the tip in the distal superior vena cava, but not in the right atrium.

All access devices must be managed by staff trained in aseptic procedures following written protocols. This is the only way to ensure the prevention of dangerous and avoidable complications such as catheter-related sepsis.

Nutrients

The nutrients are compounded in a multilayer plastic bag under sterile conditions in the pharmacy. The needs of the majority of patients can be met from a range of standard nutrition solutions. Commonly used parenteral solutions provide 9 or 14 g of nitrogen as amino acids, and 1600 or 2200 non-protein calories as glucose and lipid in 2.5–3.0 litres of water, with 80 mmol of sodium and 60 mmol of potassium. Calcium, magnesium and phosphate are also included, as are commercial preparations of vitamins and trace elements. These provide the currently recommended daily requirements of micronutrients. The need for additional supplements should be considered in patients who are severely depleted when parenteral nutrition is initiated, and in those patients who do not receive a nutrient bag every day. Some patients, especially those with Crohn disease, have suffered from selenium depletion. Conversely, it used to be thought that more manganese is absorbed from the diet than is the case. This led to the formulation of preparations that supplied too much manganese for intravenous administration. Such problems are more common in patients who receive parenteral nutrition at home.

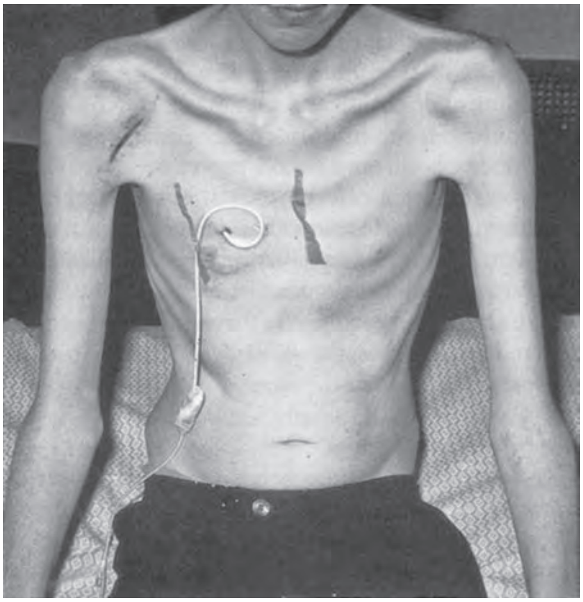


Figure 12.4 Tunnelled central catheter.

The most common adjustments to the nutrient prescription include fluid and electrolytes in patients with increased losses, or reduced tolerance with renal impairment. Considerations of lipid stability limit the addition of electrolytes, especially divalent ions. A separate electrolyte infusion is sometimes needed.

Nutrient delivery

The nutrients are delivered via an infusion pump. As with enteral nutrition, cyclical feeding is preferred in the stable patient, but continuous infusion may be required in the stressed patient. Cyclical infusions facilitate mobilization, and are associated with less fluid retention and fat deposition. Catheter patency is maintained with a heparin lock when the patient is disconnected from the infusion. All catheter procedures must be undertaken according to strict aseptic protocols to minimize the risk of complications.

Complications

Complications of parenteral nutrition may be considered in three groups:

1. nutritional and metabolic
2. catheter related
3. effect on other organ systems.

Nutritional and metabolic complications

Fluid overload, hyperglycaemia and electrolyte imbalance are common potential problems, especially in unstable patients. Patients who are severely malnourished may suffer from the refeeding syndrome (see above). Additional electrolytes may need to be given to some patients; this applies particularly to phosphate, potassium and magnesium. Stressed patients are prone to hyperglycaemia so some of the energy should be provided as lipid and insulin may be needed. Rebound hypoglycaemia can occur after discontinuing the infusion of concentrated glucose solutions until endogenous insulin levels fall. Tapering the rate of infusion before disconnection minimizes this problem.

Long-term patients who receive little nutrition from other sources are at risk of micronutrient deficiencies. Whereas the profile of commercially available trace element solutions has been changed to cover previous deficiencies such as selenium, it is worth remembering that amounts delivered in each nutrition bag are designed to meet estimated daily needs. Thus, patients with initial depletion or who do not require a nutrition bag every day (this applies to many of the patients who receive parenteral nutrition at home) are at risk of deficiency and monitoring is important. The excessive accumulation of aluminium from protein hydrolysates and manganese from incorrectly formulated trace element solutions, leading to bone and neurological disease, respectively, should no longer occur.

The overprovision of macronutrients such as glucose and amino acids is harmful. Excess glucose will lead to hepatic steatosis and is accompanied by increased respiratory demands and the stimulation of the sympathetic nervous system. The excessive administration of amino acids has been associated with cholestasis in paediatric practice and possible alterations in bone metabolism.

Catheter-related complications

Complications associated with catheter insertion are common. They include pneumothorax, damage to adjacent vascular structures and incorrect positioning of the catheter tip. Placement with the aid of screening is important and the use of ultrasound is helpful in minimizing these risks. After correct placement there are four important catheter-related complications:

1. infection
2. thrombosis
3. occlusion
4. fracture.

Catheter-related infection

Three types of infection can be recognized: exit-site infection, tunnel infection and catheter-related septicaemia.

Exit-site infections usually respond to local dressing complemented by systemic antibiotic therapy. However, clearance will not be successful if there is a Dacron cuff, which lies adjacent to the exit site (rather than halfway along the tunnel) when it becomes contaminated. *Infection of the tunnel*, which can complicate exit-site infection, is characterized by pain and redness; it does not respond to treatment and is an indication for catheter removal. Rarely, tunnel infections may arise from retrograde spread of infection from an infected central vein. This is a very serious problem, which reflects previous inadequately treated catheter-related sepsis.

Catheter-related septicaemia is the most serious infection and can be fatal. In the short term, microbes may gain access from the exit site, but after 2 weeks the hub is the most common site of contamination. The problem is usually heralded by pyrexia and rigors during infusion. Diagnosis can be difficult, especially in patients with other sources of sepsis; there may also be other reasons for pyrexia, and these include venous thrombosis. When this complication is suspected, catheter removal is the safest course of action. However, if venous access is difficult and the need for continuing central parenteral nutrition is anticipated, drawback and peripheral blood cultures should be obtained. Semiquantitative cultures have been used to implicate the line as the primary source of infection; however, the use of the endoluminal brush promises more precise and reliable information. While awaiting the results of culture, the central catheter should be locked and the patient is given systemic antimicrobial cover against the most likely infecting organisms; vancomycin + gentamicin is a suitable regimen in the short term. Catheter salvage is possible using antibiotic and urokinase locks. There are insufficient data to recommend specific guidelines in relation to the dose of urokinase and type of antibiotic. One approach involves the insertion of a 2000 U/mL urokinase lock for 1 hour before the infusion of antibiotics; pending the results of blood cultures, vancomycin and gentamicin administered twice a day cover likely pathogens, although blood antibiotic concentrations should be monitored. The antibiotic prescription may need to change according to microbiological results. Salvage should not be attempted when *Candida* and *Staphylococcus aureus* infections are identified: it is unlikely to succeed with subcutaneous ports, and the potential risk of recrudescence and metastatic infection, including endocarditis and osteomyelitis, should be considered.

The emphasis must be on prevention. The development of catheter care protocols and management by nutrition support teams has almost eliminated the problem. The use of antimicrobial agents as additives in the nutrient solution cannot be recommended for routine practice and requires further study, but may help in the management of the difficult patient who is prone to recurrent infection.

Central vein thrombosis

Central vein thrombosis tends to develop after treatment for a few weeks (Figure 12.5). It may be heralded by pyrexia, pulmonary embolism, subclavian vein thrombosis or occlusion of the superior vena cava with facial swelling. This is a serious problem, which impairs venous access and can prove fatal. The diagnosis should be confirmed by bilateral upper limb venography. Rarely in adult practice does the thrombus involve the heart; however, if any new cardiac murmur is heard transoesophageal echocardiography is recommended. Thrombosis accompanies the use of concentrated glucose regimens and a proximal location of the catheter tip in the superior vena cava; thrombotic tendency is reduced by the administration of lipid emulsions. Patients with coagulation disorders such as antithrombin III deficiency and patients with Crohn disease are more likely to develop this problem.

Treatment with thrombolytic drugs such as streptokinase is effective in restoring venous patency. Streptokinase is infused for 48 hours, when repeat venography will usually demonstrate recanalization. The patient is then given heparin and ultimately warfarin. There is a theoretical risk of inducing pulmonary embolism from fragmented thrombi. The use of heparin alone does not lead to significant recanalization in the short term.

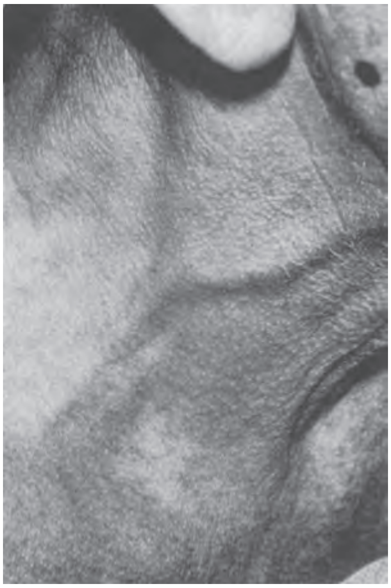


Figure 12.5 Venous thrombosis associated with central parenteral nutrition.

Many authorities use heparin in the nutrient solution as a prophylactic measure. Evidence suggests that there is a need for at least 3 units of heparin/mL of feed solution, although the possible effect on the stability of the emulsion must be considered. There is insufficient evidence to support the routine use of heparin in all patients. Low-dose warfarin may be an effective alternative, notwithstanding interactions between warfarin and the vitamin K derivatives, which occur in the lipid emulsions. The need for prophylaxis will depend on individual circumstances. For example, prophylaxis will be required in patients with a thrombotic disorder and in patients who have already developed venous thrombosis. Conversely, anticoagulation is unwise in some patients with active inflammatory disease in whom there is a significant risk of bleeding.

Catheter occlusion

Catheter occlusion may be due to kinking or luminal deposition of fibrin, lipid sludge or amorphous debris. This is a particular problem with lipid mixes, the tendency to which may be reduced by the use of an ethanol flush before the application of the heparin lock.

When occlusion occurs a chest radiograph should be obtained. Providing the position of the catheter is satisfactory, attempts at line salvage can be considered. When there is evidence for fibrin occlusion, after blood flashback or with a fibrin sleeve, urokinase can be used to clear the catheter. Streptokinase is reserved for the management of serious central vein thrombosis to avoid sensitizing the patient. When occlusion develops with the use of lipid mixes, an ethanol lock may restore the patency of an incompletely occluded catheter.

Catheter damage

The use of connection devices or extension sets reduces the need to clamp the catheter and prolongs catheter life. Repair kits are available if fracture occurs. Subcutaneous ports eventually leak. This becomes apparent when the patient complains of pain around the infusion site. New ports can be attached to the existing catheter.

Effect of parenteral nutrition on other organ systems

Parenteral nutrition may affect the hepatobiliary system, the immune system and the skeleton. Any theoretical detrimental effect must be weighed against the known impairment of organ function associated with the malnutrition that this treatment seeks to prevent or reverse.

Changes in organ function may be a reflection of the underlying disease, malnutrition, drug treatment and lack of oral or enteral nutrition, as well as the effect of parenteral nutrition. The development of hepatobiliary disease in patients who are treated with prolonged parenteral nutrition illustrates this point. Some forms of hepatobiliary disease occur more commonly in patients with inflammatory bowel disease; the lack of oral nutrition leads to biliary sludge and possible changes in intestinal permeability with toxin absorption. The administration of excessive amounts of glucose may cause hepatic steatosis and in neonatal patients the excessive prescription of amino acids has been incriminated in the development of cholestasis. Metabolic bone disease in some patients is attributable to previous malnutrition and exposure to corticosteroid therapy. The influence of parenteral nutrition on bone metabolism is controversial.

Catheter complications can cause serious disease in other systems. Endocarditis and osteomyelitis have been described following catheter-related sepsis, and pulmonary embolism can complicate central vein thrombosis. With due care and the observance of good written protocols parenteral nutrition is relatively safe, and is life saving for a significant group of patients with short- or long-term intestinal failure.

GUIDE TO FURTHER READING

Edington J, Kon P, Martin CN. *Prevalence of malnutrition in patients in general practice. Clin Nutr* 1996;**15**:60–63.

European Nutrition for Health Alliance. *Malnutrition.* London, UK: ENHA, 2008. Available from: european-nutrition.org

Green CJ. Existence, causes and consequences of hospital and community disease related malnutrition and clinical and financial benefits of nutritional intervention. *Clin Nutr 1999;***18***(Suppl.* 2):3–28.

Heatley RV, Green JH, Losowsky MS (eds). *Consensus in Clinical Nutrition.* Cambridge, UK: Cambridge University Press, 1994.

Jensen GL, Mirtallo J, Compher C, *et al.* Adult starvation and disease-related malnutrition: a proposal for etiology based diagnosis in the clinical practice setting from the International Consensus Guideline Committee. *J Parenter Enteral Nutr* 2010;**34**:156–159.

Payne JJ, Grimble G, Silk D (eds). *Artificial Nutrition Support in Clinical Practice.* London, UK: Edward Arnold, 1995.

Pennington CR. Disease and malnutrition in British hospitals. *Proc Nutr Soc* 1997;**56**:393–407.

Pirlich M, Schütz T, Kemps M, *et al.* Prevalence of malnutrition in hospitalized medical patients: impact of underlying disease. *Dig Dis* 2003;**21**:245–251.

Rombeau JL, Caldwell MD. *Clinical Nutrition: Parenteral Nutrition*, 2nd edn. Philadelphia, PA: W.B. Saunders, 1993.

Rombeau JL, Rolandelli RH (eds). *Clinical Nutrition: Enteral and Tube Feeding.* Philadelphia, PA: W.B. Saunders, 1997.

CHAPTER 13

Interventional radiology in general surgery

MOHAMAD S. HAMADY

Percutaneous biopsy

Over the past several decades, percutaneous image-guided biopsy has become one of the most common interventional radiological procedures. In particular, rapid advances in imaging technology, biopsy instruments and techniques have made once difficult-to-reach lesions, historically requiring surgical intervention, accessible and amenable to the ambulatory setting. Furthermore, the development of novel histopathology methods with a greater sensitivity and specificity have increased diagnostic yield.

Biopsy instruments

Biopsy instruments can be classified into aspiration and cutting core needles (Figure 13.1). Each technique has its particular advantages in obtaining either cells for cytological assessment or tissue for histopathological examination.

Aspiration needles

Aspiration needles are predominantly used to procure tissue for cytological analysis. They are particularly useful for the safe biopsy of small lesions (<2.5 cm diameter) lying next to vital organs or vascular structures and have an associated complication rate of <1%. The needles themselves are designed with a bevel-shaped end and range in size from 18 to 25 G. Overall, the aspiration technique is recognized to have a high sensitivity for cancer diagnosis but a low cellular accuracy and cancer specificity.

Cutting needles

Cutting biopsy needles are designed to obtain segments of tissue for histological examination rather than individual cells. The cutting mechanism of the instrument is further subclassified as end-cutting (predominantly for solid lesion biopsy) or side-cutting (mainly for soft lesion biopsy).

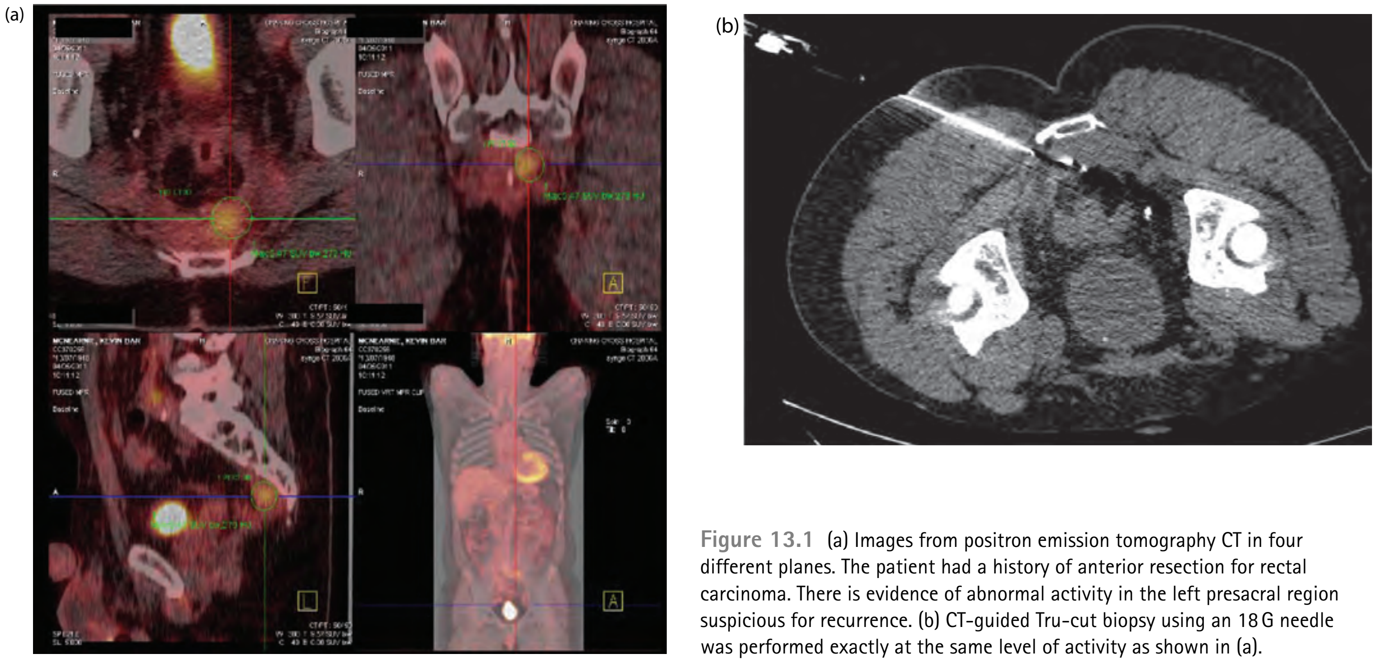


Figure 13.1 (a) Images from positron emission tomography CT in four different planes. The patient had a history of anterior resection for rectal carcinoma. There is evidence of abnormal activity in the left presacral region suspicious for recurrence. (b) CT-guided Tru-cut biopsy using an 18 G needle was performed exactly at the same level of activity as shown in (a).

Cutting needles are advantageous in being able to provide a larger tissue yield and therefore a higher chance of a more accurate diagnosis, particularly for lesions where the primary origin is unknown following standard radiological assessment. They have a low complication profile and are especially safe when used in conjunction with image guidance for percutaneous procedures.

Percutaneous abscess drainage

Indications

Over the course of the past three decades, percutaneous abscess drainage has displaced open surgical intervention to become the first-line treatment for evacuation of abdominal and pelvic collections. In addition to being a minimally invasive technique that obviates the need for open surgery and general anaesthesia, the percutaneous approach allows for the majority of abscesses to be drained fully and safely without the associated morbidity and mortality inherent with laparotomy, while saving the surgical option for particularly resistant or extensive abscesses.

The development of modern, high-resolution imaging techniques coupled with the growing number of experienced interventional radiologists has meant that the vast majority of fluid collections and abscesses in the peritoneal cavity, retroperitoneal space and even, in certain cases, specific organs are today amenable to percutaneous drainage.

Fluid drainage is absolutely indicated when there is a clinical or laboratory suspicion that a defined collection is infected, or causing a mass organ effect. The aim of such drainage is either definitive treatment or as a diagnostic adjunct in preparation for definitive surgery. Patients with diverticulitis and associated abscess collection are occasionally treated with percutaneous drainage in preparation for bowel resection. Additionally, when patients are being palliated or are otherwise unsuitable for surgical intervention, percutaneous drainage offers a minimally invasive treatment option.

Indications for percutaneous drainage include:

- to obtain a fluid sample for diagnostic investigations
- treatment of abscess through a non-surgical modality
- in preparation for a definitive surgical procedure
- as a palliative measure in patients not suitable for surgical intervention.

Contraindications

Although there are no absolute contraindications to percutaneous drainage, the following are considered to be the main relative contraindications to employing this minimally invasive approach:

- severe coagulopathy
- cardiopulmonary compromise
- absence of a safe entry route to the drainage site
- complex, multiloculated collection
- pregnancy
- patient co-operation.

The ultimate decision on whether to attempt a percutaneous drainage or intervene surgically in these circumstances should be taken following multidisciplinary discussion involving surgeons, intensivists and interventional radiologists.

Complications

Major complications such as severe bleeding (due to uncorrected coagulopathy or vascular or solid organ injury), sepsis and organ injury necessitating surgical intervention are relatively rare, with an estimated incidence of <10%. Several measures may be undertaken periprocedurally to reduce these risks:

- correction of coagulation defects
- group and save and transfuse individuals with a low haemoglobin count
- careful imaging of complex or multisegmented, multilocation collections
- early recognition of pathology that requires a surgical intervention, e.g. tumour, complex abscess connecting to an organ structure
- preplanning the safest route of entry with suitable pre- and intraprocedure imaging with consideration of using different windowing techniques of scanners and patient position to obtain the best views and access
- ensuring that the expertise of the interventional radiologist matches the requirement of the procedure to be carried out; if the complexity of the procedure requires greater expertise then consider referring the patient to a specialist centre
- have a surgical team available to assess the patient in case of any complications.

As we have already discussed, the success of drainage through an interventional approach depends not only on the skill and expertise of the interventional radiologist but equally, if not more importantly, on careful consideration of the patient and their underlying pathology, a detailed understanding of the anatomy of the abdomen as a whole and the area to be drained, especially in cases of complex, multiloculated collections. Finally, it is equally essential to recognize when the interventional approach will not succeed and surgery is the only treatment.

Overall, the success rate for percutaneous abscess drainage approaches 90%. However, this figure falls significantly in the presence of multiloculated, large, thick-walled collections having a complex anatomy, in which successful drainage may be achieved in as little as 30% of cases.

Management of the *in situ* percutaneous drain

Although most patients tolerate drain insertion very well, it is not uncommon for some to experience a transient postprocedure fever in the first 24 hours following catheter placement. Most patients will recover within 48 hours but deterioration or lack of improvement in clinical condition may necessitate further patient review and investigation to ensure that there is no new pathology or indeed worsening of existing pathology.

Following insertion of the drainage catheter, its position and patency should be confirmed and checked on a regular basis for the duration that it remains *in situ*. Many radiologists prefer to use double-lumen catheters for drainage purposes.

These provide two advantages. First, the presence of a second lumen permits closed system irrigation with ejection of infused saline (equal to the volume injected) and pus from the cavity, giving an indication of both catheter position and patency. Furthermore, flushing the catheter regularly each day helps to keep the catheter unblocked and clear of debris or pus that can settle there. Second, contrast material can be injected intracorporeally and the patient imaged with fluoroscopy or CT to obtain a further indication of the catheter position, size of the cavity being drained and the volume that remains before deciding on whether the catheter can be removed.

For most cases, routine abdominal imaging to check drainage adequacy is not necessary. Clinical improvement coupled with a reduced drain output to <30 mL/day (in the absence of a blocked drain) is generally an indication that the job of the drain has been completed satisfactorily. Conversely, the persistence of a high output can suggest a fistulous communication being present intracorporeally and it is in this situation that further imaging with (water soluble) contrast enhancement through the drain can prove invaluable.

In addition to the general percutaneous management of intra-abdominal abscesses, it is worth considering how collections associated with individual organs may need a more specialized approach to be adequately treated.

Management of organ-specific abscess and collections

Liver

Liver abscesses can be classified according to their nature as pyogenic, amoebic or fungal or alternatively by their number as singular or multiple. Pyogenic abscesses are mostly polymicrobial and secondary to recent surgical intervention, infected tumour, infected biliary obstruction or bacteraemia in immunocompromised patients. Singular abscesses usually occur on the right lobe and, although the causative agent is often unknown, it can be associated with *Klebsiella pneumoniae*. Multiple abscesses, on the other hand, are usually bilobular and biliary in origin, with *Escherichia coli* being the most commonly isolated organism.

Diagnosis is most commonly made on clinical symptoms and confirmed by radiological imaging such as ultrasound or CT. For the most part, treatment is predominantly medical with supportive and appropriate antimicrobial therapy with percutaneous drainage being reserved for non-responsive and large (>8 cm) abscesses (Figure 13.2).

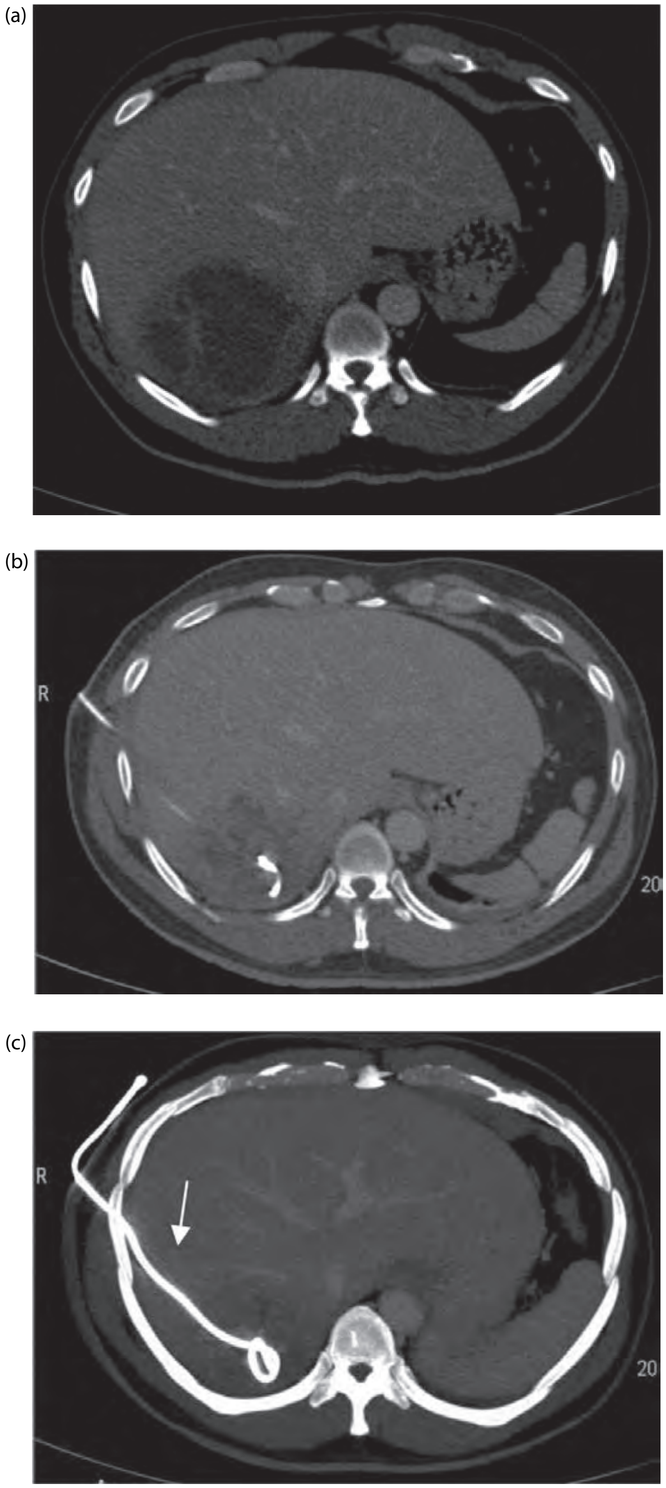


Figure 13.2 (a) Enhanced CT scan slice through the liver shows an 8 cm abscess in segment VI. (b) Enhanced CT slice following percutaneous CT-guided drain insertion. Note the significant reduction in abscess size. (c) Maximum intensity projection image of the liver similar to Figure 13.4b. The safe intercostal path of the drain is shown (white arrow).

Pancreas

The pancreas (Figure 13.3) can be subject to formation of a large variety of collections which may be amenable to percutaneous drainage. Severe acute pancreatitis in particular can result in phlegmon formation, pseudocyst and haemorrhagic necrosis. Pseudocysts and necrotic tissues can be further complicated by added infection. Additionally, up to 40% of patients may develop fluid collections as the pancreatitis resolves. Definitive differentiation of these pathologies is on the basis of imaging findings complemented by clinical features, timeline and aspiration.

Pseudocysts typically develop between 4 and 6 weeks after an episode of severe acute pancreatitis and show discrete wall

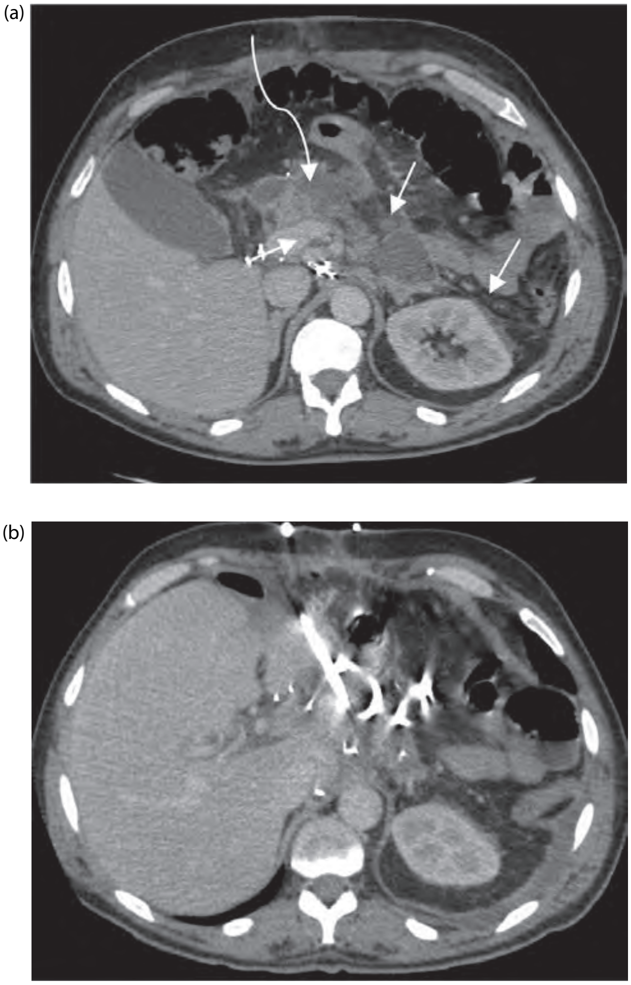


Figure 13.3 (a) CT scan of a 51 year old man following gastrectomy at the level of the pancreas. There is a multiloculated fluid collection around the pancreas (arrows). Note the tight safe window of drainage (curved arrow). (b) CT scan following insertion of two double-lumen pigtail drains under CT guidance. There is complete resolution of fluid collection.

enhancement on CT scan. It has been suggested that serum amylase reflects the maturity of the pseudocyst. As with fluid collections, the great majority of pseudocysts resolve spontaneously and require no radiological or surgical intervention. However, intervention may be indicated in the presence of superimposed infection or mass effects such as recurrent pain and bowel obstruction.

Pancreatic necrosis appears as areas of non-enhancement in contrast-enhanced CT. Treatment of infected pancreatic necrosis has traditionally been the domain of the general surgeon. However, in modern practice, percutaneous image-guided drainage with vigorous irrigation of the necrotic pancreas can achieve good results when performed under meticulous aseptic technique and antibiotic cover. There is a growing trend of draining infected necrosis using large-bore catheters (up to 30 F) with surgical therapy being reserved only for intractable or extensive cases of tissue necrosis.

Renal and perirenal

Renal abscesses are commonly the result of haematogenous spread of infection or pyelonephritis associated with renal stones, obstructive hydronephrosis and septicaemia. Renal abscesses may rupture into the collecting system or into the perinephric space. Both sites of abscess collection are suitable for percutaneous drainage, with surgical stent insertion, e.g. nephrostomy and ureteric stents being necessary to treat the underlying condition.

Bowel

Enteric abscess can complicate any inflammatory or infective bowel process or result from iatrogenic injury during surgical procedures (Figures 13.4–13.6). Diverticular disease, appendicitis and inflammatory bowel disease are most frequently associated with bowel abscess formation. The main role of percutaneous drainage is temporizing the patient's status to facilitate definitive surgery on an elective basis. The exception to this general rule is in diverticular-related abscess, in which drainage occasionally has a role in treating cases with perforation limited to the mesentery or pelvis without significant faecal contamination.

Periappendiceal abscess may be secondary to perforated/gangrenous appendicitis or ongoing infection after appendicectomy. When a large appendix abscess is present, the preferred management is to treat the patient conservatively by inserting a drainage catheter under CT guidance and perform definitive elective surgery at a later date. In some, limited, instances, drainage may even obviate the need for surgery. Postoperative periappendiceal abscess also responds favourably to percutaneous drainage.

Crohn's abscess is associated with a high level of morbidity and can occur secondary to direct extension from the involved bowel segment, haematogenous spread to a distal site or as a result of anastomotic leak following surgery. Percutaneous drainage may improve morbidity, but, particularly in cases of anastomotic leak, the catheter drainage can continue for an extended period of time until the anastomotic line is completely healed.

Gastrointestinal balloon dilatation and stenting

Oesophagus

Image-guided intervention plays an important role in the management of a wide range of oesophageal pathology (Figure 13.7). A significant number of patients present with non-resectable oesophageal cancer. Palliative treatment therefore has a significant role in the management of advanced oesophageal carcinoma with the aim of reducing the effect of tumour bulk. A wide range of palliative measures including thermal ablation, photodynamic therapy, radiotherapy and oesophageal stenting (including plastic and metal, covered or uncovered stents) are available.

The indications for oesophageal stenting in the setting of malignancy are:

- intrinsic oesophageal compression causing significant dysphagia
- tracheo-oesophageal fistula: a spontaneous malignant fistula can occur secondary to local tumour invasion; since malignant fistulas will

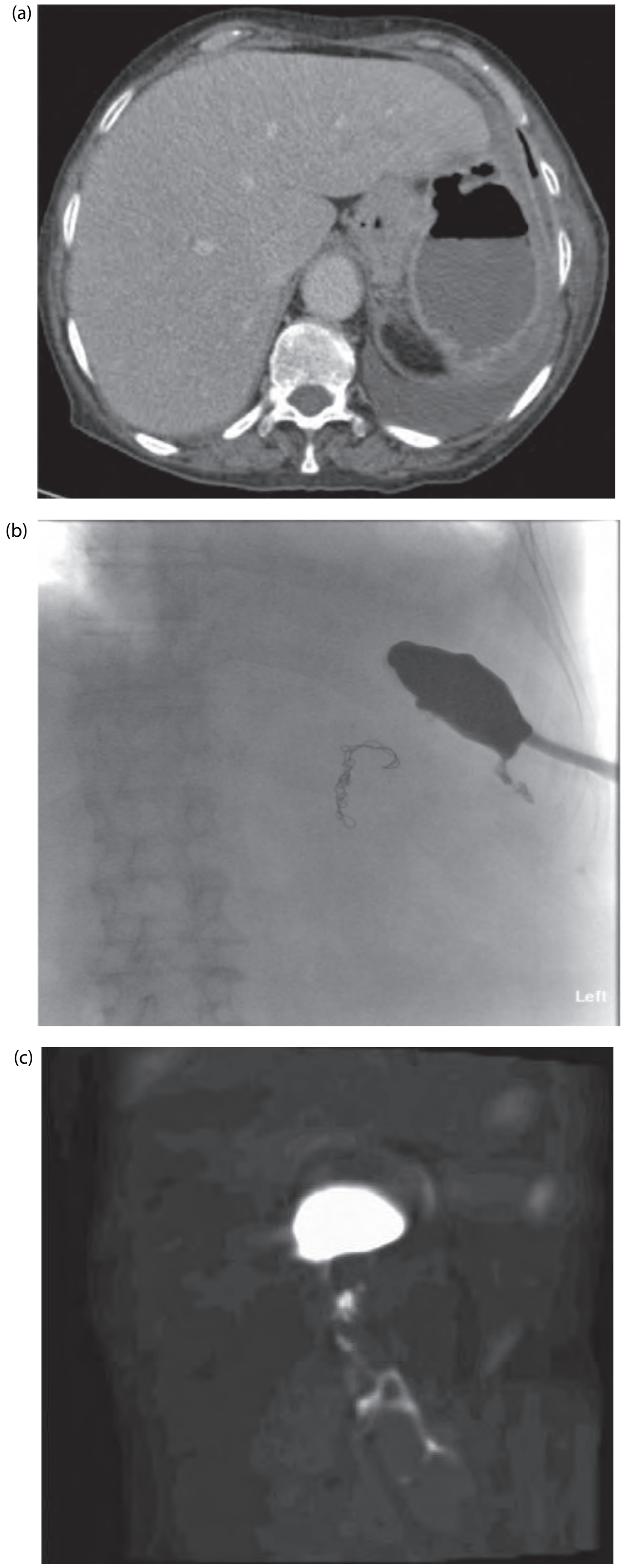


Figure 13.4 (a) CT scan of the abdomen at the level of the liver. There is an abscess collection containing gas in the left subphrenic space. (b) Linogram through the existing drain in the left subphrenic space. It is difficult to establish bowel communication with the abscess cavity. (c) Three-dimensional image obtained from fluoroscopy shows definite communication between the abscess cavity and the descending colon.

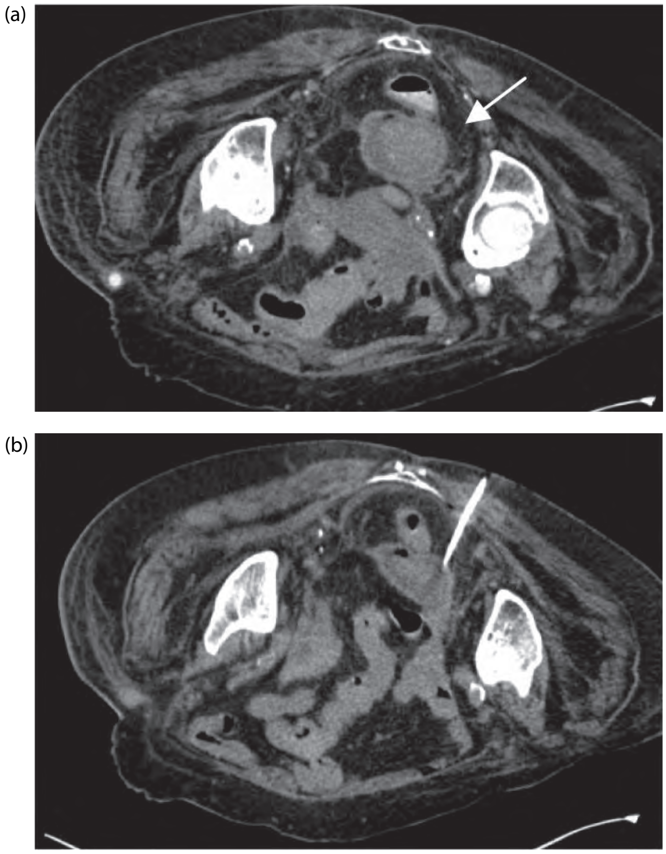


Figure 13.5 (a) Axial CT slice at the sacrococcygeal level shows a prerectal abscess collection (arrow). (b) Axial CT slice at the same level as in (a) shows a drain *in situ* with significant abscess resolution.

not heal spontaneously, self-expandable covered metallic stents are the treatment of choice in those cases

- primary or secondary mediastinal tumours resulting in extrinsic compression
- anastomotic tumour recurrence not responding to balloon dilatation.

Anastomotic stricture and pyloric dysfunction following oesophagectomy and gastric pull-up (Figure 13.8) are relatively common complications, ranging between 27% and 42% and 12% and 13%, respectively. Early diagnosis and treatment is crucial to avoid anastomotic leak, dysphagia, malnutrition and poor quality of life. Balloon dilatation is effective in treating benign stricture and pyloric dysfunction. However, malignant stricture responds better to stenting. In a recent study that involved treating 48 patients with anastomotic stricture after oesophagectomy, radiologically guided balloon dilatation was effective in 98% of benign strictures versus 64% of malignant strictures. In the same study, balloon dilatation was effective in 92% of early pyloric dysfunction but less effective in delayed pyloric dysfunction (63%). In another study that involved 269 patients, anastomotic stricture was noticed in 42% of cases. Endoscopic dilatation was successful in 78% of patients after a median of three dilatation sessions.

In cases of oesophageal leak and fistula, the approach is slightly different. The majority of those cases are secondary to surgery

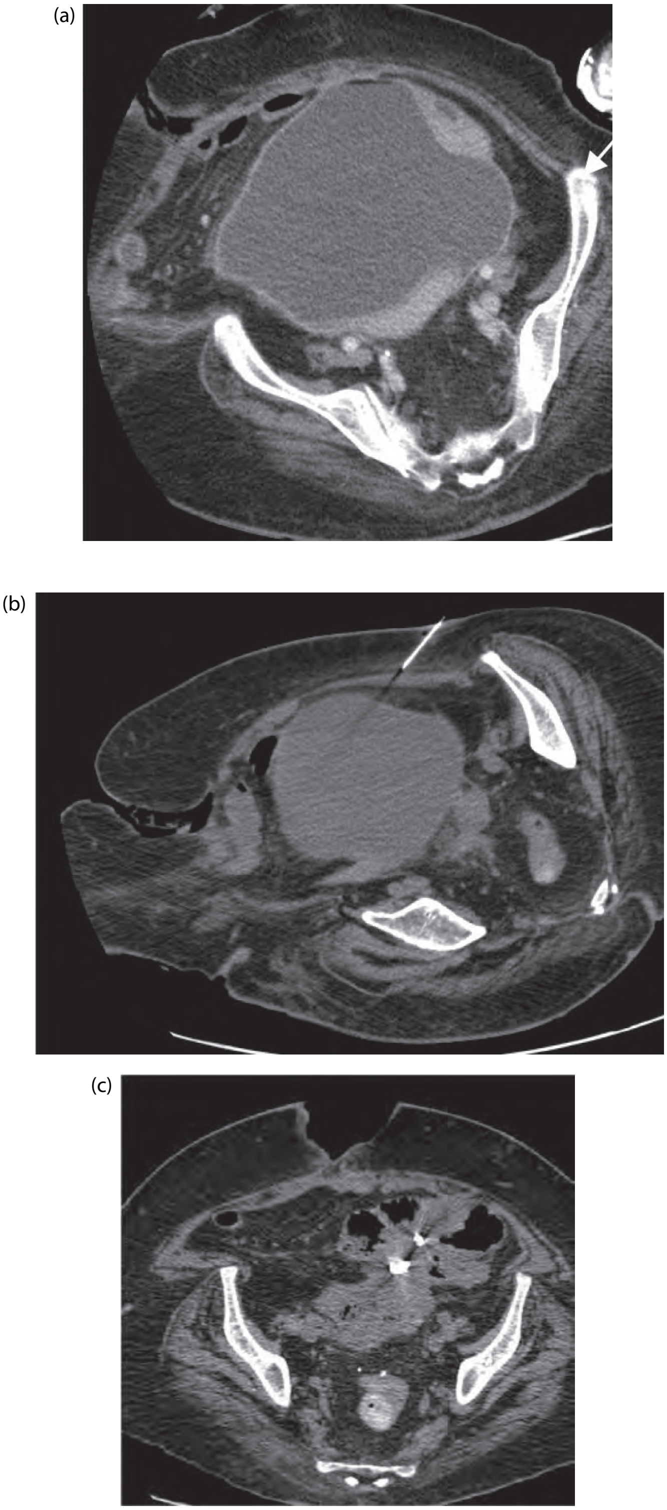


Figure 13.6 (a) Axial CT slice at the level of the pelvis shows a large abscess collection following sigmoidectomy (arrow). (b) Axial CT slice shows the path of image-guided needle insertion through a safe window. (c) Axial CT slice after drain insertion, showing complete resolution of the collection.

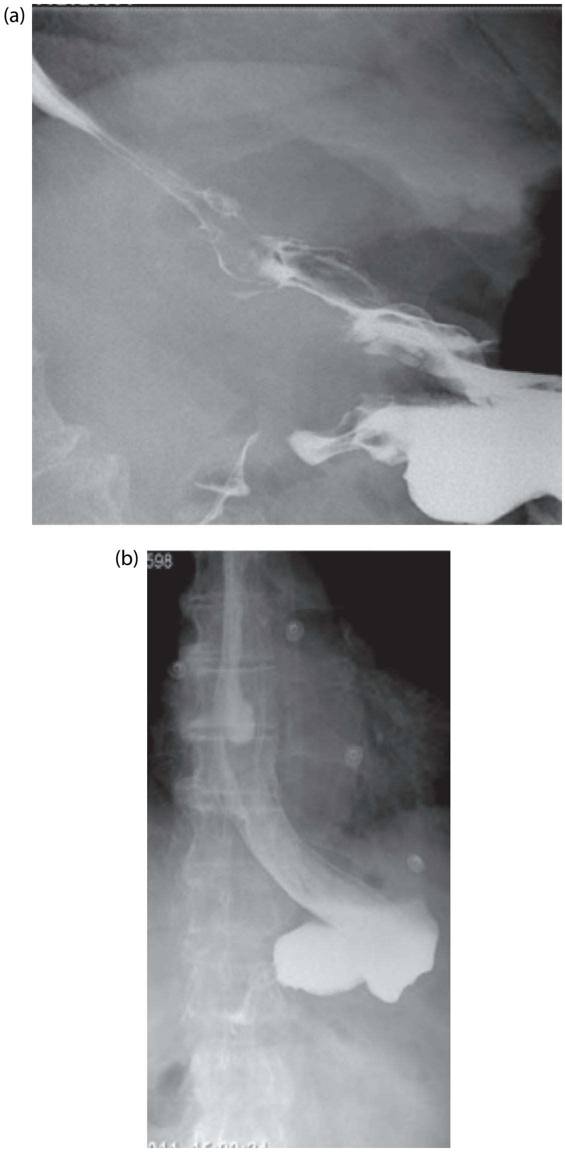


Figure 13.7 (a) Patient with advanced oesophageal carcinoma and grade 3 dysphagia. Barium swallow shows stricture and irregularity of the distal third of the oesophagus and gastro-oesophageal junction. (b) Gastrograffin swallow following oesophageal stenting. There is good flow of contrast through the stent with no hold-up of contrast.

or underlying malignancy. In iatrogenic cases, such as those seen following oesophagectomy and gastric pull-up, the pyloric function must be restored using balloon dilatation if needed. The patient should be nil by mouth and a nasogastric tube inserted to the level above the pylorus to avoid excessive dilatation of the pouch. Most of these patients respond well to these conservative measures. Rarely, repeat dilatation of the pylorus or even stenting of the pylorus could be needed to enhance drainage.

The relative contraindications for oesophageal stenting include recent radiotherapy (in the last 4 weeks), tumour higher than the cricopharyngeal level and severe tracheal compression that could be exacerbated by oesophageal stenting.

The technical success rate of radiologically placed oesophageal stents in improving dysphagia ranges between 90% and 98%. Although stenting has no impact on the survival rate of those patients with advanced disease, it can substantially improve the quality of life.

The complications of radiologically guided oesophageal stenting are generally low. The main complications include migration, perforation, bleeding, chest pain and reflux.

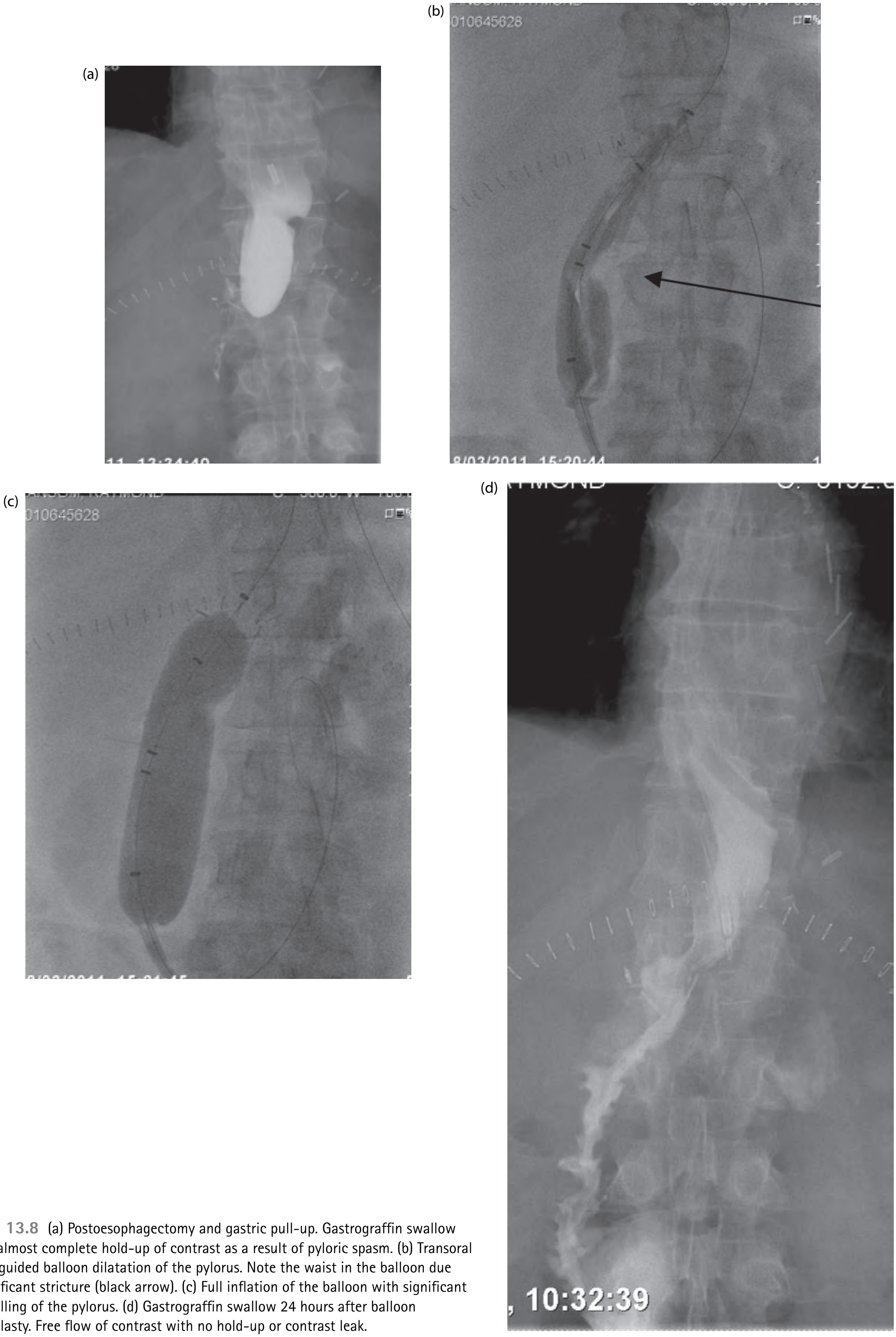


Figure 13.8 (a) Postoesophagectomy and gastric pull-up. Gastrograffin swallow shows almost complete hold-up of contrast as a result of pyloric spasm. (b) Transoral image-guided balloon dilatation of the pylorus. Note the waist in the balloon due to significant stricture (black arrow). (c) Full inflation of the balloon with significant remodelling of the pylorus. (d) Gastrograffin swallow 24 hours after balloon pyloroplasty. Free flow of contrast with no hold-up or contrast leak.

Bleeding can occur in 3–8% of cases. It is generally self-limited. In cases of persistent or heavy bleeding arterial embolization is indicated. Migration can be seen in 0–6% of cases depending on the type of stent, accurate placement and site of the lesion. Self-expandable covered stents are associated with a low incidence of migration. Stents extending into the stomach have a slightly higher chance of migration than stents covering middle oesophageal lesions. Once stent migration happens, restenting using telescopic stents is indicated. The migrated stent can be left *in situ*, if it is causing no abdominal pain or obstruction, or can be removed through a small gastrostomy incision or using endoscopic forceps. Initial chest pain or discomfort is expected in most patients. However, more severe or persistent pain is seen in higher lesions or when large-diameter stents are used. Reflux can be observed in patients with stents crossing the cardia. Antacid and proton pump inhibitors are indicated in those cases.

Once the stent is deployed and the operator is satisfied with its position, the patient is advised to remain nil by mouth for a few hours before resuming clear fluid. The patient should be instructed to start eating well-chewed food, avoid swallowing large pieces of food and drink carbonated drink during the meal to avoid blockage of the stent.

Colon stenting and balloon dilatation

Acute large bowel obstruction (Figure 13.9) can be secondary to a variety of malignant and benign conditions. Colorectal carcinoma is the most common cause of acute large bowel obstruction. Other less frequent malignant causes include metastatic infiltration or infiltration from adjacent malignancy. Benign causes of large bowel obstruction include inflammatory bowel disease and diverticulitis or postradiation stricture.

Patients with acute obstruction have a high mortality and morbidity. The traditional treatment of acute large bowel obstruction is open surgery, usually involving multiple stages. However, surgical intervention is associated with significant mortality and morbidity. In a case–control study comparing the mortality and morbidity between emergency and elective surgery for colon cancer, the mortality was reduced from 34% to 7% and morbidity was reduced from 64% to 24%. The use of metallic stents as a palliative or temporizing measure prior to definitive surgery has become accepted clinical practice. The indications for colorectal stenting are:

- decompression of malignant bowel obstruction to facilitate elective surgery
- palliative decompression of malignant bowel obstruction that is deemed unresectable
- preoperative decompression of benign bowel stricture such as diverticular disease or following radiotherapy
- rarely, in the management of a colonic fistula.

The technical success of stent placement in the colon ranges from 79% to 100%. In a systematic review involving 10 studies and a total of 451 patients comparing colonic stent and open surgery in patients with colon carcinoma, stenting was successful in 92% of cases and was associated with a shorter hospital stay and fewer complications. In a review of 234 patients who had a colonic stent placed for malignant colonic obstruction, the clinical success rate ranged from 75% to 100% with a mean of 90%.

Complications are relatively rare and consist of migration, perforation, bleeding and pain. The overall rate of complication is around 14%. Following successful colonic stent placement, vital signs should be monitored and a plain abdominal film is usually obtained 24–48 hours following the procedure. Stent position and adequate stent expansion are the two important features to look for in the plain film. If stent migration is detected a watchful waiting policy can be taken, as the majority of stents will pass per rectum. Surgery or endoscopic removal is considered if the patient reobstructs. If the stent is inserted for palliative reasons, a stool softener may be helpful to reduce stent obstruction.

Balloon dilatation is reserved for benign strictures such as Crohn's disease and anastomotic strictures. Most patients require one balloon dilatation and the majority of cases respond favourably with a low complication and recurrence rate.

Gastrointestinal bleeding

Lower gastrointestinal bleeding

Lower gastrointestinal bleeding (Figure 13.10) occurs at a rate of approximately 25 cases per 100 000 population per year, with a higher reported incidence among males. Diverticular disease and angiodysplasia are the most common causes of bleeding followed by haemorrhoids, polyps and ischaemic enteritis.

Sigmoidoscopy and colonoscopy represent the first line of investigation and management. The drawback of scoping is the trouble in identifying the exact source of bleeding because of difficulty in visualizing the source of haemorrhage during an active bleed.

Radiological methods such as radioisotopes, angiography and multidetector CT scan allow bleeding points to be identified more readily. Scintigraphy is commonly performed using sulphur colloid or ^{99m}Tc-labelled red blood cells. ^{99m}Tc in particular allows image acquisition over a longer period of time, which can be especially useful for the identification of intermittent bleeding (at rates as low as 0.1–0.5 mL/min) and can increase the sensitivity of subsequent diagnostic angiography. However, the sensitivity and specificity of scintigraphy is highly variable ranging between 24% and 91%.

Percutaneous angiography is performed via a puncture in the femoral or brachial artery. Positive angiography needs active bleeding to occur at a minimum rate of 0.1 mL/min and has sensitivity between 43% and 87%. It is used to demonstrate contrast extravasation, tumour vascularity and/or early draining vein (indicative of angiodysplasia).

With the recent advances in CT technology, there is growing evidence that CT scanning can be helpful in diagnosing the bleeding site, provides information about the underlying pathology, e.g. tumour, and provides information about the arterial anatomy, aiding a quicker and safer angiography. The reported sensitivity, in small series, is around 77%. The limitations of CT scan are the need to give 120 mL of iodinated contrast, radiation and the presence of active bleeding at a rate of 0.5 mL/min.

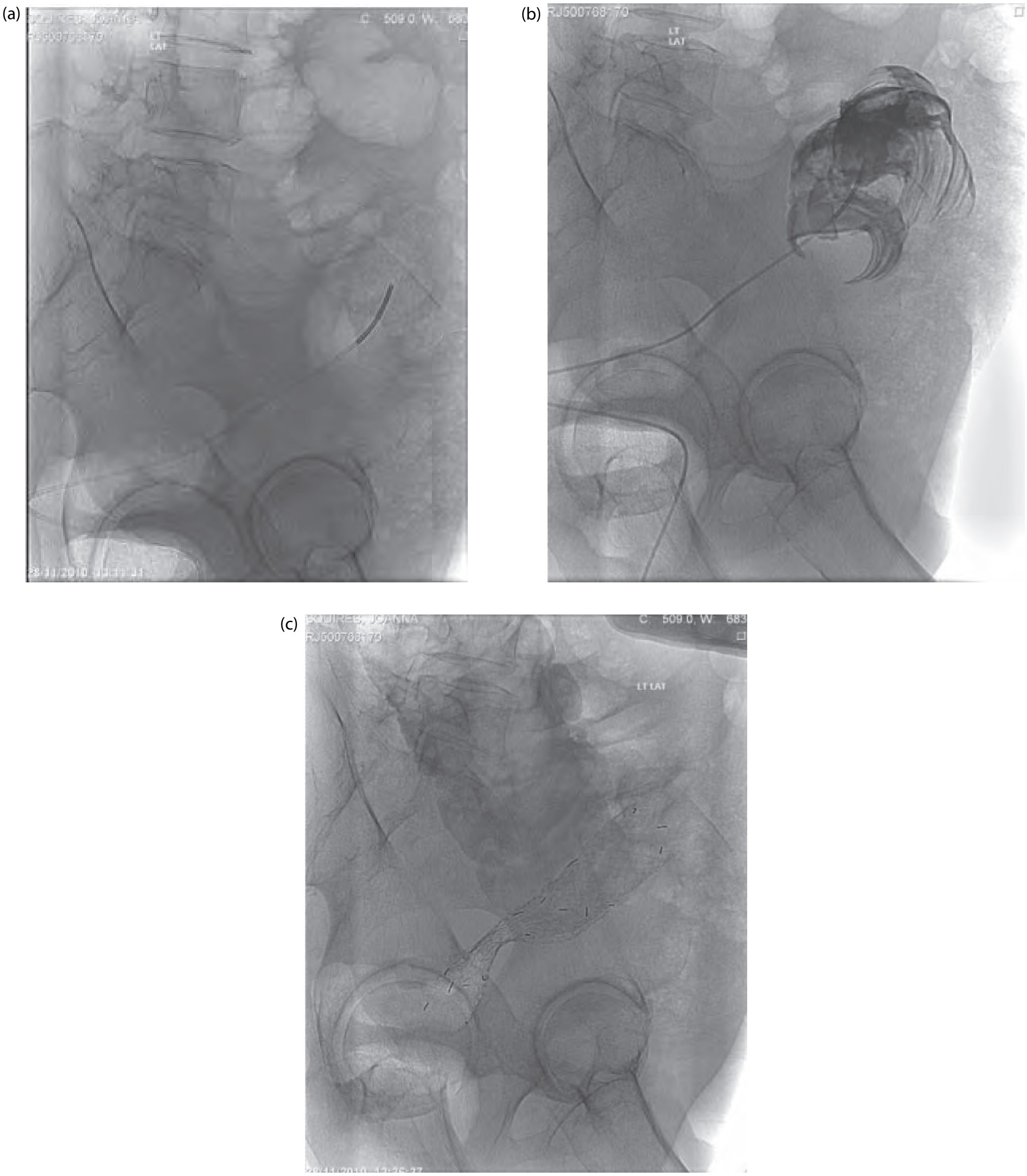


Figure 13.9 (a, b) An 81 year old woman with advanced rectal carcinoma and large bowel obstruction. Lateral fluoroscopy shows that the catheter has crossed the lesion following wire manipulation. (c) A self-expandable stent was deployed across the lesion to relieve the obstruction.

To ensure positive angiography, the patient should be actively bleeding. Bleeding per rectum is probably not the best predictor of active bleeding since the colon has a large capacity to hold blood. Moreover, bleeding has probably stopped when old blood is still passing per rectum. The best clinical indicators of the likelihood of successful angiography are high pulse rate (>100 beats/min), low blood pressure (<90 mmHg) and the need for blood transfusion (5 or more units), i.e. the shocked patient. Nevertheless, elderly patients with limited cardiovascular reserve may not tolerate massive bleeding, with a drastic effect on end-organ perfusion. Therefore, the threshold of performing angiography in this subset of patients should be low. In cases in which active bleeding is uncertain or in patients with chronic low-volume bleeding with negative colonoscopy, angiography is also indicated.

There is no absolute contraindication to visceral angiography, but the relative contraindications include:

- contrast allergy
- renal failure
- coagulopathy.

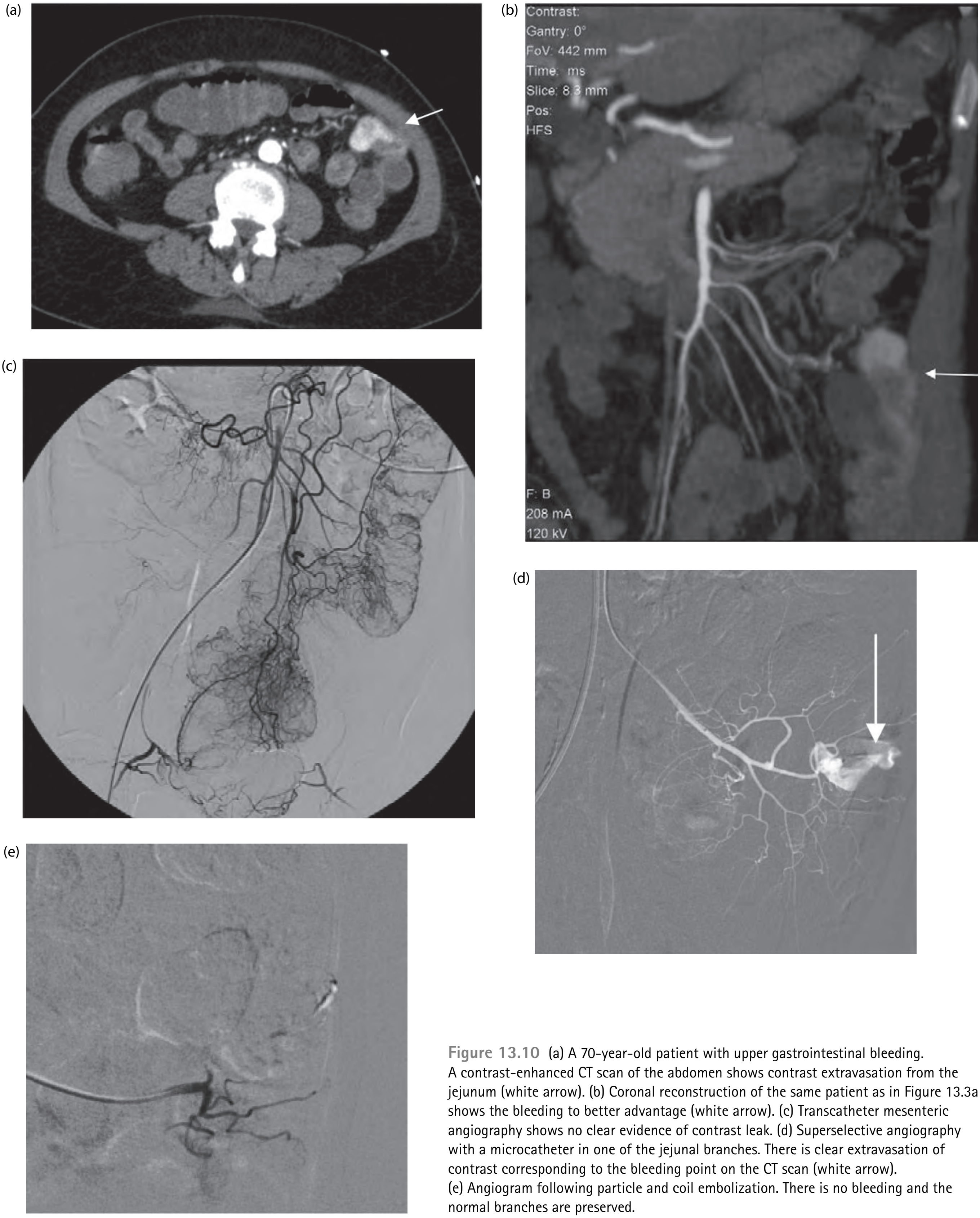


Figure 13.10 (a) A 70-year-old patient with upper gastrointestinal bleeding. A contrast-enhanced CT scan of the abdomen shows contrast extravasation from the jejunum (white arrow). (b) Coronal reconstruction of the same patient as in Figure 13.3a shows the bleeding to better advantage (white arrow). (c) Transcatheter mesenteric angiography shows no clear evidence of contrast leak. (d) Superselective angiography with a microcatheter in one of the jejunal branches. There is clear extravasation of contrast corresponding to the bleeding point on the CT scan (white arrow). (e) Angiogram following particle and coil embolization. There is no bleeding and the normal branches are preserved.

These should be balanced against the benefits gained from angiography such as identification and sealing of the bleeding point. Prophylactic measures such as steroid premedication, adequate hydration and correcting coagulation defects or using a closure device at the puncture site can limit the aforementioned risks.

Gastrointestinal bleeding can be treated using transcatheter embolization. The rationale of this treatment is to decrease the blood flow pressure to the bleeding site to allow clot formation. This can be achieved by using 2.7–3 F microcatheters introduced coaxially through the 5 F selective catheters. This technique allows the embolization process to take place as distal and as close as possible to the extravasation point, together with minimizing the risk of infarction of an adjacent segment of normal bowel. The embolic material used for this purpose includes particles such as polyvinyl alcohol, microcoils and gelfoam particles. The reported technical success of embolization is high, ranging between 73% and 89%. The main reasons for technical failure include small target vessel size, extreme vessel tortuosity, vasospasm and lack of technical expertise. In current practice, transcatheter embolization is accepted as a safe and effective treatment modality. The main causes of rebleeding are failure to achieve the required endpoint of low flow pressure, an unappreciated underlying pathology such as tumour or inflammatory bowel disease and development of collaterals.

The main complication of gastrointestinal embolization is non-target embolization, which could result in bowel infarction or ischaemic stricture. With the advancement in catheter technology, embolic materials and embolization technique, the rate of failed embolization requiring surgical intervention has reduced to around 7%.

Following the embolization procedure, the vascular sheath should be left *in situ* if there is a need for arterial monitoring, uncorrected coagulopathy or a high chance of repeating the procedure. Otherwise, the sheath can be removed and haemostasis is achieved by manual groin compression or by a closure device to the arteriotomy puncture. The patient should be rehydrated, blood products replaced and oxygen administered to minimize the risk of low bowel perfusion and tissue oxygenation. The vital signs, blood counts and abdomen should be monitored regularly for signs of continuous bleeding and/or peritonitis and bowel infarction.

Upper gastrointestinal bleeding

The majority of non-variceal upper gastrointestinal bleeding stops spontaneously. Unlike colonic haemorrhage, endoscopy is the first line of management with adequate injection of the bleeding point producing favourable outcomes. Surgery and embolization are considered in cases that are refractory to endoscopic treatment. Embolization was first performed by Rosch in 1972 and was reserved for high-risk patients. With advancements in catheters and embolic material technology, embolization has now become the second line of management in patients with failed therapeutic endoscopy. As well as being a minimally invasive technique, embolization has the advantage of being suitable for surgically unfit patients and repeatable in cases of persistent bleeding.

Peptic ulcer disease is the most common cause of non-variceal upper gastrointestinal bleeding with the duodenum being the most common site of bleeding. Iatrogenic causes of bleeding such as sphincterotomy, percutaneous transhepatic cholangiogram and live abscess drainage are becoming more frequent.

Non-variceal upper gastrointestinal bleeding is intermittent. Identification of the bleeding site needs the patient to be actively bleeding. Colloid scintigraphy can detect bleeding as slow as 0.1 mL/min whereas red blood cell scintigraphy can detect bleeding at 0.2–0.4 mL/min. Catheter angiography using the current digital subtraction technology can detect bleeding at approximately 0.6 mL/min. Multidetector CT angiography in arterial and venous phases is a promising non-invasive imaging modality with a sensitivity similar to colloid scintigraphy. The advantages of CT scan include repeatability and high sensitivity (which helps in detecting low-flow bleeding together with the ability to plan the next step in management). In cases of failed endoscopy, angiography may need to be considered as an alternative to surgery without the need to wait for full resuscitation as this measure can be taken in the interventional radiology suite. The role of angiography is thus both diagnostic and therapeutic.

There is no absolute contraindication to formal angiography. However, careful management of the access site in patients with severe coagulopathy is important. The sheath could be left in the artery until coagulopathy is corrected with or without the use of occlusion material to seal the puncture hole.

Angiography can detect 40–60% of cases of active non-variceal upper gastrointestinal bleeding. Embolization can control angiographically detected bleeding in 90% of cases with overall clinical success in 50–90% of patients. In a retrospective study of 163 patients with non-variceal upper gastrointestinal bleeding treated by embolization, the total mortality was 33%. Walsh *et al.* found a 17% mortality rate following successful embolization versus 62% if embolization fails.

Specific complications of embolization are rare. Duodenal stricture, ulceration and necrosis are recognized complications. The incidence of complication is higher in patients with a history of previous gastric surgery, radiotherapy or advanced atherosclerosis. Splenic infarction is another rare complication if embolization of the splenic artery is encountered. However, splenectomy in a stable patient is a relatively straightforward surgical procedure should it be required. Liver infarction is expected if there is portal vein thrombosis. Therefore, great care should exercised by the operator in those circumstances.

Postembolization follow-up should continue for at least 30 days to detect late duodenal stricture. If duodenal stricture is expected clinically, radiological confirmation is then advocated with a view to treating symptomatic patients.

Chronic mesenteric ischaemia

Chronic mesenteric ischaemia is a rare condition that usually affects people older than 60 years. Women are typically three times more affected than men. The symptoms of chronic mesenteric ischaemia are usually present when two out of three vessels (coeliac trunk, superior mesenteric and inferior mesenteric arteries) are significantly stenosed (Figure 13.11).

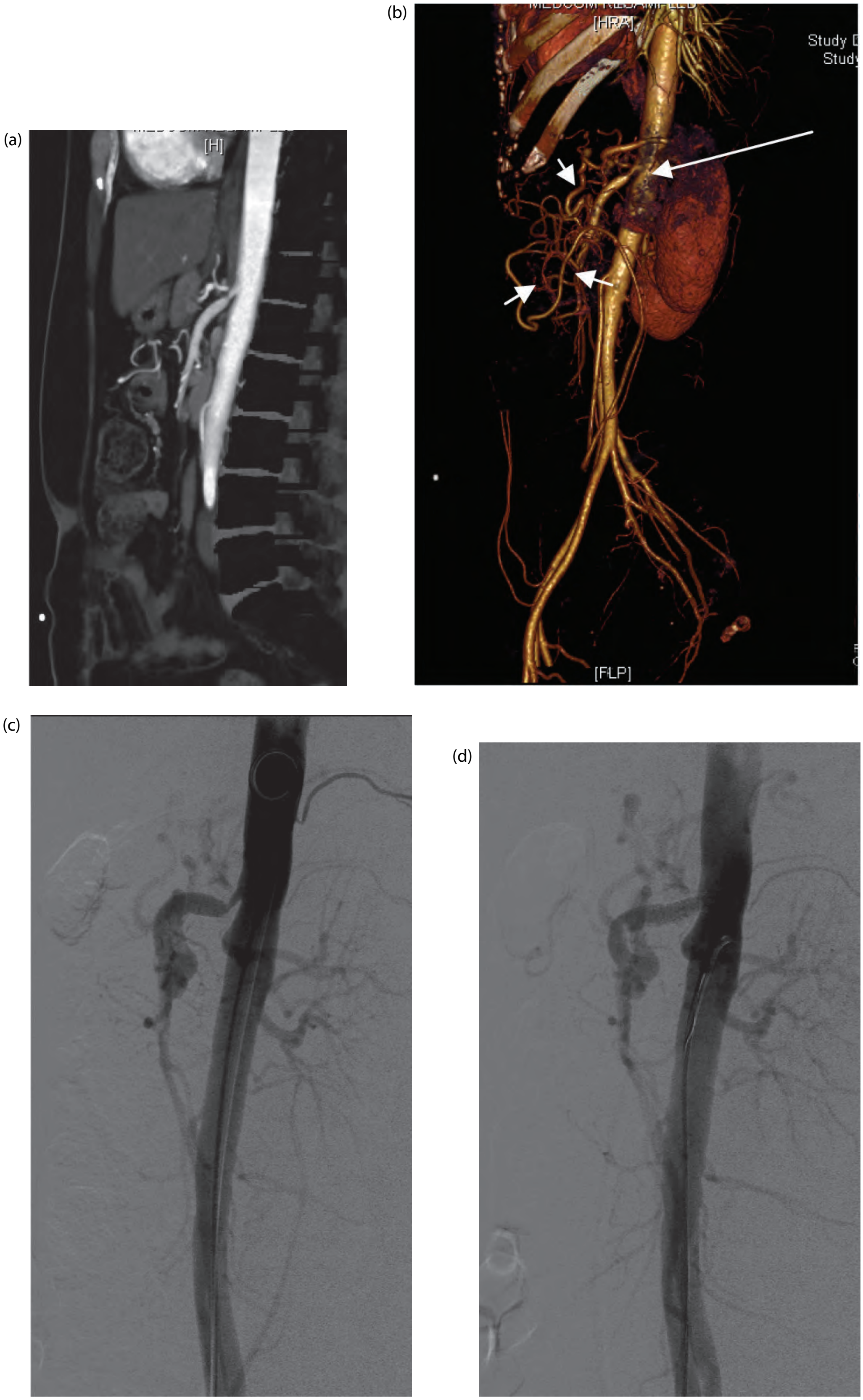


Figure 13.11 (a) Sagittal maximum intensity projection CT angiography shows severe stenosis of the superior mesenteric artery and occluded coeliac trunk at its origin. (b) Volume-rendering CT angiography image shows the stenosis of the superior mesenteric artery (long arrow) and visceral collaterals (short arrows). (c) Catheter angiography confirms the severe stenosis at the origin of the superior mesenteric artery. (d) Catheter angiography following stent angioplasty. Note complete resolution of the stenosis.

However, single-vessel stenosis can sometimes be enough to cause ischaemic symptoms.

Atherosclerosis is the most common cause of mesenteric vessel stenosis. Other causes include Takaysu disease, fibromuscular dysplasia, thrombangitis obliterans and radiotherapy. The majority of patients with visceral artery disease have associated cardiovascular morbidity in other vascular beds.

Clinical diagnosis

Clinically, patients present with postprandial abdominal pain, associated with weight loss, nausea and occasional diarrhoea (which is attributed to villous atrophy). The diagnosis is frequently delayed as weight loss provokes extensive investigations for malignancy in the elderly population, whom it typically affects.

Radiological diagnosis

There are several imaging modalities available to establish the diagnosis of chronic mesenteric ischaemia. CT angiography (CTA) is the most helpful modality. It provides information about the severity of the stenosis, the extent of the disease and assesses the state of the entire aorta as well as the bowel wall. Moreover, it helps the operator to plan the best approach to tackle the lesion in question. CTA is, also, helpful to assess the postoperative complications of open or endovascular treatment. However, CTA with contrast can be nephrotoxic with serious effects on patients with pre-existing poor kidney function.

Magnetic resonance angiography (MRA) is a good alternative to CTA (safe for patients with impaired kidney function) and can provide valuable information regarding the status of vessels and grafts as well as the bowel vascularity. Nevertheless, MRA is inferior to CTA in assessing the patency of stents and presence of atheromatous plaques. Colour Doppler ultrasound has also been used to diagnose visceral artery occlusion or stenosis but this modality depends on operator experience and is known for its poor resolution when compared with CTA or MRA.

Treatment

Patient selection for any intervention is crucial. All non-vascular causes for abdominal pain and weight loss should be excluded both clinically and radiologically prior to intervention. Patients with two or more vessel stenoses can proceed to intervention. Asymptomatic patients with two severely diseased or occluded vessels and awaiting major intestinal surgery may additionally benefit from having at least one vessel revascularized.

Open surgery is considered the traditional treatment of chronic mesenteric ischaemia. Open surgery includes embolectomy and thrombectomy, endarterectomy and bypass grafts. The success rate in relieving symptoms is quite high and 5 year patency rates are reported to be between 75% and 89%. However, this approach is associated with significant morbidity and mortality in comorbid patients. The reported mortality rate ranges from 4% to 17% and morbidity rate from 20% to 33%.

The advances in interventional radiological techniques paved the way for a valid and successful alternative approach to bypass surgery. Percutaneous angioplasty and stenting is a minimally invasive approach with high technical and clinical success. The minimally invasive nature of the procedure allows for early mobility and recovery as well as reduced inpatient stay. The procedure can be performed under local anaesthesia via a femoral or transbrachial approach. Since the first reported case of mesenteric angioplasty by Furrer *et al.* in 1980, several studies have shown good mid- and long-term results of visceral artery stenting and angioplasty. The technical success rate, especially in recent publications, is very high (approaching 100%), reflecting the advances in technology and maturation of operators' experience.

Razavi and Chung (2004) reported their experience in 70 symptomatic patients with 97% technical success and 99% clinical success. The restenosis rate over 3 years was 10%.

In a systematic review of 16 studies including 328 patients, the technical success was 91% and most failures were secondary to the presence of the arcuate median ligament. The reintervention rate for restenosis was 27%, resulting in good primary and secondary assisted patency. However, the authors were unable to analyse the secondary intervention data owing to irregularity in reporting the relevant data. Nevertheless, the review confirmed a very low 30 day mortality rate of 3%.

GUIDE TO FURTHER READING

Aabakken L. Nonvariceal upper gastrointestinal bleeding. *Endoscopy* 2001;**33**:16–23.

Abbas SM, Bissett IP, Holden A, *et al.* Clinical variables associated with positive angiographic localization of lower gastrointestinal bleeding. *ANZ J Surg* 2005;**75**:953–957.

AbuRahma AF, Stone PA, Bates MC, *et al.* Angioplasty/stenting of the superior mesenteric artery and celiac trunk: early and late outcomes. *J Endovasc Ther* 2003;**10**:1046–1053.

Adam A, Ellul J, Watkinson AF, *et al.* Palliation of inoperable esophageal carcinoma: a prospective randomized trial of laser therapy and stent placement. *Radiology* 1997;**202**:344–348.

Aina R, Oliva VL, Therasse E, *et al.* Arterial embolotherapy for upper gastrointestinal hemorrhage: outcome assessment. *J Vasc Interv Radiol* 2001;**12**:195–200.

Alavi A, Dann RW, Baum S, Biery DN. Scintigraphic detection of acute gastrointestinal bleeding. *Radiology* 1977;**124**:753–756.

Allen RC, Martin GH, Rees CR, *et al.* Mesenteric angioplasty in the treatment of chronic intestinal ischemia. *J Vasc Surg* 1996;**24**:415–423.

Allison D, Hemmingway A, Cunningham D. Angiography in gastrointestinal bleeding. *Lancet* 1982;**2**:30e3.

Balasegaram M. Management of hepatic abscess. *Curr Probl Surg* 1981;**81:**3.

Bandi R, Shetty PC, Sharma RP, *et al.* Superselective arterial embolization for the treatment of lower gastrointestinal hemorrhage. *J Vasc Interv Radiol* 2001;**12**:1399–1405.

Baron TH, Dean PH, Yates MR, *et al.* Expandable metal stents for the treatment of colonic obstruction: techniques and outcomes. *Gastrointest Endosc* 1998;**47**:277–286.

Binkert CA, Ledermann HP, Jost R, *et al.* Acute colonic obstruction: clinical aspects and cost-effectiveness of preoperative and palliative treatment with self-expanding metallic stents–a preliminary report. *Radiology* 1998;**206**:199–204.

Butler JA, Huang J, Wilson SE. Repeated laparotomy for post-operative intra-abdominal sepsis: an analysis of outcome factors. *Arch Surg* 1987:**122:**702–706.

Campbell DM, Geraghty JG, McBride K, *et al.* Radiologically controlled balloon dilatation of rectal strictures. *Clin Radiol* 1991;**44**:77–78.

Choo IW, Do YS, Suh SW, *et al.* Malignant colorectal obstruction: treatment with a flexible covered stent. *Radiology* 1998;**206**:415–421.

Chou F, Sheen-Chen SM, Chen YS, Chen MC. Single and multiple pyogenic liver abscesses: clinical course, etiology, and results of treatment. *World J Surg* 1997;**21**:384–389.

Cinat ME, Wilson SE, Din AM. Determinants for successful percutaneous image-guided drainage of intra-abdominal abscess. *Arch Surg* 2002;**137**:845–849.

De Gregorio MA, Mainar A, Tejero E, *et al.* Acute colorectal obstruction: stent placement for palliative treatment-results of a multicenter study. *Radiology* 1998;**209**:117–120.

De Lange, Shaffer HA. Rectal strictures: treatment with fluoroscopically guided balloon dilation. *Radiology* 1991;**178**:475–479.

d'Othée BJ, Surapaneni P, Rabkin D, *et al.* Microcoil embolization for acute lower gastrointestinal bleeding. *Cardiovasc Intervent Radiol* 2006;**29**:49–58.

Earlam R, Cunha-Melo JR. Oesophageal squamous cell carcinomas. II. A critical view of radiotherapy. *Br J Surg* 1980;**67**:457–461.

Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med* 2003;**349**:2241–2252.

Evangelista PT, Hallisey MJ. Transcatheter embolization for acute lower gastrointestinal hemorrhage. *J Vasc Interv Radiol* 2000;**11**:601–606.

Ferguson CB, Mitchell RM. Nonvariceal upper gastrointestinal bleeding: standard and new treatment. *Gastroenterol Clin North Am* 2005;**34**:607–621.

Fiorito J, Brandt L, Kozicky O, *et al.* The diagnostic yield of superior mesenteric angiography: correlation with the pattern of gastrointestinal bleeding. *AJR Am J Roentgenol* 1989;**84**:878–881.

Furrer J, Grüntzig A, Kugelmeier J, *et al.* Treatment of abdominal angina with percutaneous dilatation of an arteria mesenterica superior stenosis. Preliminary communication. *Cardiovasc Intervent Radiol* 1980;**3**:43–44.

Gazelle GS, Haaga JR. Biopsy needle characteristics. *Cardiovasc Intervent Radiol* 1991;**14**:12–16.

Gerzof SG, Robbins AH, Johnson WC, *et al.* Percutaneous catheter drainage of abdominal abscesses: a five-year experience. *N Engl J Med* 1981;**305**:653–657.

Gordon RL, Ahl KL, Kerlan RK, *et al.* Selective arterial embolization for the control of lower gastrointestinal bleeding. *Am J Surg* 1997;**174**:24–28.

Guy GE, Shetty PC, Sharma RP, *et al.* Acute lower gastrointestinal hemorrhage: treatment by superselective embolization with polyvinyl alcohol particles. *AJR Am J Roentgenol* 1992;**159**:521–526.

Hansen KJ, Wilson DB, Craven TE, *et al.* Mesenteric artery disease in the elderly. *J Vasc Surg* 2004;**40**:45–52.

Hollier LH, Bernatz PE, Pairolero PC, *et al.* Surgical management of chronic intestinal ischemia: a reappraisal. *Surgery* 1981;**90**:940–946.

Homs MY, Wahab PJ, Kuipers EJ, *et al.* Esophageal stents with antireflux valve for tumors of the distal esophagus and gastric cardia: a randomized trial. *Gastrointest Endosc* 2004;**60**:695–702.

Honkoop P, Siersema PD, Tilanus HW, *et al.* Benign anastomotic strictures after transhiatal esophagectomy and cervical esophagogastrostomy: risk factors and management. *J Thorac Cardiovasc Surg* 1996;**111**:1141–1148.

Imbembo AL. Diverticular disease of the colon. In: Sabiston D (ed.) *Textbook of Surgery*, 14th edn. Philadelphia, PA: W.B. Saunders, 1992.

Jawhari A, Kamm MA, Ong C, *et al.* Intraabdominal and pelvic abscess in Crohn's disease: results of non-invasive and surgical treatment. *Br J Surg* 1998;**85:**367–371.

Jimenez JG, Huber TS, Ozaki CK, *et al.* Durability of antegrade synthetic aortomesenteric bypass for chronic mesenteric ischemia. *J Vasc Surg* 2002;**35**:1078–1084.

Johannsen EC, Sifri CD, Madoff LC. Pyogenic liver abscess. *Infect Clin North Am* 2000;**14**3:547–563.

Johnson WC, Gerzof SG, Robbins AH, Nabseth DC. Treatment of abdominal abscesses: comparative evaluation of operative drainage versus percutaneous catheter drainage guided by computed tomography or ultrasound. *Ann Surg* 1981;**194**:510–520.

Johnston KW, Lindsay TF, Walker PM, *et al.* Mesenteric arterial bypass grafts: early and late results and suggested surgical approach for chronic and acute mesenteric ischemia. *Surgery* 1995;**118**:1–7.

Kasirajan K, O'Hara PJ, Gray BH, *et al.* Chronic mesenteric ischemia: open surgery versus percutaneous angioplasty and stenting. *J Vasc Surg* 2001;**33**:63–71.

Khot UP, Lang AW, Murali K, Parker MC. Systematic review of the efficacy and safety of colorectal stents. *Br J Surg* 2002;**89**:1096–1102.

Kougias P, El Sayed HF, Zhou W, Lin PH. Management of chronic mesenteric ischemia. The role of endovascular therapy. *J Endovasc Ther* 2007;**14**:395–405.

Koval G, Benner KG, Rösch J, Kozak BE. Aggressive angiographic diagnosis in acute lower gastrointestinal hemorrhage. *Dig Dis Sci* 1987;**32**:248–253.

Kuhle W, Sheiman G. Detection of active colonic hemorrhage with use of helical CT: findings in a swine model. *Radiology* 2003;**228**:743.

Kuo WT, Lee DE, Saad WE, *et al.* Superselective microcoil embolization for the treatment of lower gastrointestinal hemorrhage. *J Vasc Interv Radiol* 2003;**14**:1503–1509.

Lambiase RE, Deyoe L, Cronan JJ, Dorfman GS. Percutaneous drainage of 335 consecutive abscesses: results of primary drainage with 1-year follow-up. *Radiology* 1992;**184**:167–179.

Lang EK, Springer RM, Glorioso LW 3rd, Cammarata CA. Abdominal abscess drainage under radiologic guidance: causes of failure. *Radiology* 1986;**159**:329–336.

Lee SH. The role of oesophageal stenting in the non-surgical management of oesophageal strictures. *Br J Radiol* 2001;**74**:891–900.

Longstreth GF. Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. *Am J Gastroenterol* 1997;**92**:419–424.

Luketich JD, Christie NA, Buenaventura PO, *et al*. Endoscopic photodynamic therapy for obstructing esophageal cancer: 77 cases over a 2-year period. *Surg Endosc* 2000;**14**:653–657.

Lundell L. Upper gastrointestinal hemorrhage: surgical aspects. *Dig Dis* 2003;**21**:16–18.

Mainar A, Tiger E, Maynar M, *et al*. Colorectal obstruction: treatment with metallic stents. *Radiology* 1996;**198**:761–764.

Maspes F, Mazzetti di Pietralata G, Gandini R, *et al*. Percutaneous transluminal angioplasty in the treatment of chronic mesenteric ischemia: results and 3 years of follow-up in 23 patients. *Abdom Imaging* 1998;**23**:358–363.

Mauro MA, Koehler RE, Baron TH. Advances in gastrointestinal intervention: the treatment of gastroduodenal and colorectal obstructions with metallic stents. *Radiology* 2000;**215**:659–669.

McLarty AJ, Deschamps C, Trastek V, *et al*. Esophageal resection for cancer of the esophagus: long-term function and quality of life. *Ann Thorac Surg* 1997;**63**:1568–1572.

McLean GK, Cooper GS, Hartz WH, *et al*. Radiologically guided balloon dilation of gastrointestinal strictures. Part 2. Results of long-term follow-up. *Radiology* 1987;**165**:41–43.

Montgomery RS, Wilson SE. Intraabdominal abscess: image-guided diagnosis and therapy. *Clin Infect Dis* 1996;**23:**28–36.

Morgan R, Adam A. Use of metallic stents and balloons in the esophagus and gastrointestinal tract. *J Vasc Interv Radiol* 2001;**12**:283–297.

Mueller PR, vanSonnenberg E, Ferrucci JT Jr. Percutaneous drainage of 250 abdominal abscesses and fluid collections. Part II. Current procedural concepts. *Radiology* 1984;**151**:343–347.

Mueller PR, Saini S, Wittenburg J, *et al*. Sigmoid diverticular abscesses: percutaneous drainage as an adjunct to surgical resection in 24 cases. *Radiology* 1987;**164:**321–325.

Mueller PR, White EM, Glass-Royal M *et al*. Infected abdominal tumors: percutaneous catheter drainage. *Radiology* 1989;**173**:627–629.

Neff R. Pancreatic pseudocysts and fluid collections: percutaneous approaches. *Surg Clin North Am* 2001;**81:**399–403.

Nicholson AA, Ettles DF, Hartley JE, *et al*. Transcatheter coil embolotherapy: a safe and effective option for major colonic haemorrhage. *Gut* 1998;**43**:79–84.

Norwood SH, Civetta JM. Abdominal CT scanning in critically ill surgical patients. *Ann Surg* 1985;**202:**166–175.

Nunez D, Huber JS, Yrizzaery JM, *et al*. Nonsurgical drainage of appendiceal abscesses. *AJR Am J Roentgenol* 1986;**146:**587–589.

O'Donnell CA, Fullarton GM, Watt E, *et al*. Randomized clinical trial comparing self-expanding metallic stents with plastic endoprostheses in the palliation of oesophageal cancer. *Br J Surg* 2002;**89**:985–992.

Okuda K, Tanikawa K, Emura T, *et al*. Nonsurgical percutaneous transhepatic cholangiography-diagnostic significance in medical problems of liver. *Am J Dig Dis* 1974;**19**:21–36.

Orringer MB, Marshall B, Chang AC, *et al*. Two thousand transhiatal oesophagectomies: changing trends, lessons learned. *Ann Surg* 2007;**246**:363–374.

Paul DL, Bloom GP. Appendiceal abscess. *Arch Surg* 1982;**117:** 1017–1019.

Peck DJ, McLoughlin RF, Hughson MN, Rankin RN. Percutaneous embolotherapy of lower gastrointestinal hemorrhage. *J Vasc Interv Radiol* 1998;**9**:747–751.

Pisani P, Parkin DM, Bray F, Ferlay J. Estimates of the worldwide mortality from 25 cancers in 1990. *Int J Cancer* 1999;**83**:18–29. Erratum in: *Int J Cancer* 1999;**83**:870–873.

Razavi M, Chung HH. Endovascular management of chronic mesenteric ischemia. *Tech Vasc Interv Radiol* 2004;**7**:155–159.

Rives KK, Harty JL, Amin M. Renal abscess: emerging concepts of diagnosis and treatment. *J Urol* 1980;**124**:444–446

Rösch J, Dotter CT, Brown MJ. Selective arterial embolization. A new method for control of acute gastrointestinal bleeding. *Radiology* 1972;**102**:303–306.

Rose SC, Quigley TM, Raker EJ. Revascularization for chronic mesenteric ischemia: comparison of operative arterial bypass grafting and percutaneous transluminal angioplasty. *J Vasc Interv Radiol* 1995;**6**:339–349.

Sabharwal T, Hamady MS, Chui S, *et al*. A randomised prospective comparison of the Flamingo Wallstent and Ultraflex stent for palliation of dysphagia associated with lower third oesophageal carcinoma. *Gut* 2003;**52**:922–926.

Sabharwal T, Gulati MS, Fotiadis N, *et al*. Randomised comparison of the FerX Ella antireflux stent and the ultraflex stent: proton pump inhibitor combination for prevention of post-stent reflux in patients with esophageal carcinoma involving the esophago-gastric junction. *J Gastroenterol Hepatol* 2008;**23**:723–728.

Schenker MP, Duszak R Jr, Soulen MC, *et al*. Upper gastrointestinal hemorrhage and transcatheter embolotherapy: clinical and technical factors impacting success and survival. *J Vasc Interv Radiol* 2001;**12**:1263–1271.

Sharafuddin MJ, Olson CH, Sun S, *et al*. Endovascular treatment of celiac and mesenteric arteries stenoses: applications and results. *J Vasc Surg* 2003;**38**:692–698.

Sharma MP, Ahuja V. Management of amebic and pyogenic liver abscess. *Indian J Gastroenterol* 2001;**20**(Suppl. 1):C33–C36.

Shih MC, Angle JF, Leung DA. CTA and MRA in mesenteric ischemia. Part 2. Normal findings and complications after surgical and endovascular treatment. *AJR Am J Roentgenol* 2007;**188**:462–471.

Silver S, Weinstein A, Cooperman A. Changes in the pathogenesis and detection of intrahepatic abscess. *Am J Surg* 1979;**137:**608–610.

Smothers L, Hynan L, Fleming J, *et al*. Emergency surgery for colon carcinoma. *Dis Colon Rectum* 2003;**46**:24–30.

Steinmetz E, Tatou E, Favier-Blavoux C, *et al*. Endovascular treatment as first choice in chronic intestinal ischemia. *Ann Vasc Surg* 2002;**16**:693–699.

Sutcliffe RP, Forshaw MJ, Tandon R, *et al*. Anastomotic strictures and delayed gastric emptying after esophagectomy: incidence, risk factors and management. *Dis Esophagus* 2008;**21**:712–717.

Symonds CJ. A case of malignant stricture of the oesophagus illustrating the use of a new form of oesophageal catheter. *Trans Clin Soc Lond* 1885;**18**:155–158.

Tilney HS, Lovegrove RE, Purkayastha S *et al*. Comparison of colonic stenting and open surgery for malignant large bowel obstruction. *Surg Endosc* 2007;**21**:225–233.

vanSonnenberg E, Ferrucci JT Jr, Mueller PR *et al*. Percutaneous radiographically guided catheter drainage of abdominal abscesses. *JAMA* 1982;**247**:190–192.

vanSonnenberg E, Wittich GR, Goodacre BW, *et al*. Percutaneous abscess drainage: update. *World J Surg* 2001;**25**:362–372.

Voeller GR, Bunch G, Britt LG. Use of technetium-labeled red blood cell scintigraphy in the detection and management of gastrointestinal hemorrhage. *Surgery* 1991;**110**:799–804.

Walsh RM, Anain P, Geisinger M, *et al*. Role of angiography and embolization for massive gastroduodenal hemorrhage. *J Gastrointest Surg* 1999; **3**:61–65.

Wenger U, Johnsson E, Arnelo U, *et al*. An antireflux stent versus conventional stents for palliation of distal esophageal or cardia cancer: a randomized clinical study. *Surg Endosc* 2006;**20**:1675–1680.

Yamaguchi K, Yoshikawa K. Enhanced CT for initial localization of active lower gastrointestinal bleeding. *Abdom Imaging* 2003;**28**:634–636.

Yoon W, Jeong YY, Shin SS, *et al*. Acute massive gastrointestinal bleeding: detection and localization with arterial phase multidetector row helical CT. *Radiology* 2006;**239**:160–167.

CHAPTER 14

Trauma and the injured patient

LESLIE KOBAYASHI, RAUL COIMBRA AND DAVID HOYT

General considerations

Trauma remains one of the leading causes of death and disability throughout the world. In the USA, it is the most common cause of death in nearly all age groups under 45 (Figure 14.1). Motor vehicle accidents remain the single most common cause of trauma-related mortality, followed by gunshot wounds and falls.

The classic trimodal distribution of trauma deaths first described by Trunkey in 1983 demonstrates that almost half of all trauma deaths occur in the first hour after injury. Two additional peaks occur within the next 1–4 hours and after the first week of injury. Few patients in the first peak are likely to be saved by the healthcare system, as they die from injury before reaching medical care. In general, survival from these injuries can only be increased by preventative measures. The second peak contains approximately 30% of trauma fatalities and is generally due to haemorrhage. It is this population that constitutes the majority of preventable deaths. The goal of a 'systems' approach to trauma care is to reduce these second peak deaths by increasing efficient and reliable access to medical care following injury. A trauma system operates within a geographical region and provides for rapid transport of victims to specified hospitals within that region. These trauma centres concentrate resources and expertise for treatment of severely injured patients immediately and effectively. The third peak, occurring in days to weeks after injury, is the result of infection and multiorgan failure. Continued understanding of the epidemiology of death and immunosuppression in this peak, as well as quality improvement strategies in the intensive care unit (ICU), may be able to decrease deaths in this peak.

Biomechanics of injury

Injuries occur as a result of energy transfer; this can originate from mechanical, electric, thermal or chemical forces in the environment and results in a physiological imbalance or structural damage in the patient.

The most common type of energy in trauma is kinetic, from the motion of a vehicle, blunt object or bullet. As kinetic energy is equal to one-half mass times velocity squared, the speed of an object, rather than its weight, has the greater impact

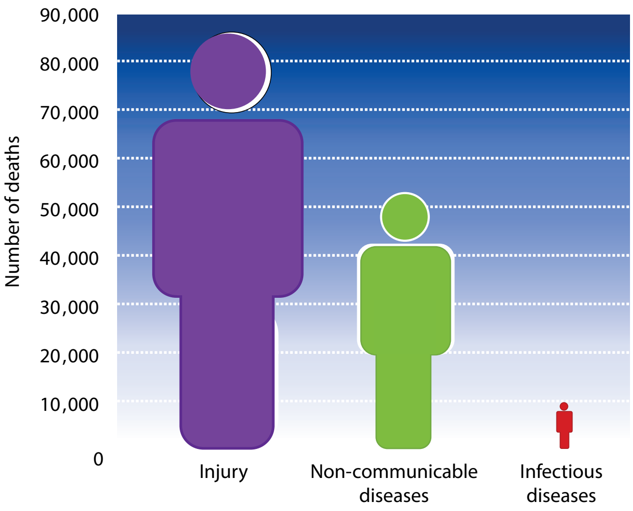


Figure 14.1 Injury deaths compared with other leading causes of death for persons aged 1–44, USA, 2007. (From Web-based Injury Statistics Query and Reporting System (WISQARS); http://www.cdc.gov/injury/wisqars.)

on the overall energy produced, and thus the damage sustained. This is well illustrated by comparing the damage produced by bullets of similar calibre but shot from low- and high-velocity weapons. When the travelling object strikes the patient, force can be imparted by direct contact, or 'impact force', or by cavitation, in which tissue moves away from the point of impact. Two types of cavitation can be created: temporary and permanent. Temporary cavities are created when the surface and structural tissues of the body are stretched but the overall shape of the body is maintained. This is common after blunt trauma. A permanent cavity is created along the lines of an impact force that causes tearing or compression of the tissues, leaving a permanent deformity of the affected tissues that can be identified later. This is often seen after penetrating trauma, in which the site of contact of the knife or bullet to the body is readily seen.

Body tissue properties such as elasticity and viscosity are also important in determining the amount of damage sustained. Relatively elastic tissue such as muscle will sustain less injury than denser, less elastic tissue such as the spleen or liver. Damage is also proportional to the rate of energy transfer, with a similar force applied over a shorter time causing more damage, again emphasizing the importance of the velocity of the objects involved. Also important is the surface area over which force is applied. Equal force applied over a smaller surface area also results in increased damage.

Blunt

Examples of blunt trauma include falls, motor vehicle collisions, pedestrians struck by vehicles, blunt assaults and sports-related injuries. Blunt trauma is the most common type of injury in the USA, with motor vehicle accidents and pedestrians struck by cars being the two most common mechanisms.

Motor vehicle accidents

Several factors influence the severity and pattern of injuries sustained after a crash. The size of the vehicle, the position of the occupant, use and type of restraint devices, the type of accident and the body habitus of the occupant all influence the injury patterns seen. When a vehicle strikes another object three collisions occur: first, the vehicle strikes the object; second, the occupant strikes the vehicle or restraint device; and, third, the internal organs strike the body cavity wall, other organs, ligaments or bone.

Frontal impact

Frontal impact occurs when the vehicle is abruptly stopped from moving forward. This rapid deceleration can result in the occupant moving up and over, or down and under the vehicle. The down and under path can result in lower extremity and pelvic fractures. The upper torso may then rotate forward, striking the vehicle, resulting in abdominal or thoracic injuries. The up and over path results in the occupant being launched over the dashboard or steering column. This results in head and facial trauma. As the torso then strikes the vehicle, compression and shear injuries can result in solid organ injuries, rupture of hollow viscous or bucket handle injury, cardiac contusion and aortic injury.

Lateral impact

In the lateral impact the vehicle is struck from the side, displacing it laterally. This can cause injury when passenger space intrusion impacts the occupant, resulting in rib, pelvic and lower extremity fractures; lung contusions; and solid organ injury. If restraints are involved the body is restrained while the head is free to continue travelling in the direction of the impact, causing flexion/rotation injuries such as spinal fracture, aortic disruption and blunt cerebrovascular injury.

Rollover and ejection

Crashes with rollover and ejection are particularly worrisome because of the potential for high-energy transfer and severe injury. During rollover the vehicle and occupant undergo multiple impacts in rapid succession, occurring at multiple different impact angles, and any type of injury is possible. Ejection also implies a high-energy transfer and rapid deceleration as the victim impacts the vehicle during exit and then impacts the ground; this can result in severe head, neck, facial and torso injuries. Victims of both ejection and rollover should be taken to the nearest trauma centre after stabilization and assumed to have life-threatening injuries until proven otherwise.

Pedestrians struck

Being struck by a vehicle is associated with very-high-energy transfer, and these patients should be considered to have multisystem trauma and transported to the nearest trauma centre for evaluation.

There are three impacts when a pedestrian is struck: first, the bumper strikes the lower extremities, the legs are displaced from under the torso; the second impact occurs when the torso then rolls up onto the bonnet of the vehicle; the third impact occurs as the victim falls away from the car. The victim can be propelled a significant distance through the air in some cases before striking the ground. Fractures of the lower extremities, pelvis, ribs and spine are common, as are injuries to the head and face. The pattern of injury will also depend on the type of vehicle, with passenger cars producing mostly lower extremity and pelvis injuries whereas trucks and vans cause more abdominal and thoracic injuries. Children are smaller and struck anatomically higher than adults with relatively few lower leg and more femur and pelvic fractures and torso injuries. Additionally, children are more likely to be dragged or fall underneath the vehicle, where they can then be run over. Significant closed head and spinal injury should be presumed until proven otherwise.

Penetrating

Penetrating trauma is categorized by the energy level imparted by the source of injury into low, medium and high velocity. Knife or ice-pick wounds are generally low energy and cause trauma only by direct laceration. Handguns and rifles can be low, medium or high energy, and cause injury by direct laceration, crushing of tissues during penetration, and temporary and permanent cavitation. The energy level is determined by the muzzle velocity of the projectile (Table 14.1). Low-energy weapons are those with muzzle velocities of less than 335 m/s. Medium-energy weapons have velocities of 335–610 m/s. High-energy weapons have velocities of greater than 610–760 m/s.

Table 14.1 Weight of projectile, muzzle velocity and approximate maximum kinetic energy of frequently used firearms

Description (calibre)	Projectile weight (g)	Muzzle velocity (feet/s)	KE (feet/lb × 10^2)
Pistols			
0.22 short	29	1000	0.5
0.38 special	158	870	1.8
9 mm Luger	125	1150	2.6
0.45	250	860	2.8
0.357 magnum	158	1430	5.0
0.44 magnum	240	1470	8.1
Rifles			
0.22 long	40	1150	0.8
5.56 mm M-16	55	3250	9.1
Winchester	170	2200	12.8

KE, kinetic energy

Higher velocity weapons tend to cause more tissue destruction because of the larger temporary cavity formed and the effects of fragmentation. Studies of high-velocity bullets demonstrate that the temporary cavity created is approximately 10–20 times the diameter of the projectile.

Common shapes of bullets include round-nose, wadcutter and semi-wadcutter. Bullets can be partially or fully jacketed, referring to the hard copper or alloy covering that will prevent fouling of the rifled barrel. Partially jacketed bullets have a soft exposed nose and are referred to as soft points. Bullets can also have hollowed out points (hollow points); both hollow and soft points will expand when striking tissue, causing tumbling, fragmentation and a wider impact zone, leading to more tissue damage. Fragmentation is a major cause of tissue destruction and occurs when the bullet breaks up after impact with the resulting pieces creating their own paths through the tissue. Alternatively, other parts of the body such as teeth or bone fragments can also become missiles, causing damage to surrounding tissues.

Shotguns are slightly different from handguns and rifles. The calibre is measured in gauge; the lower the gauge number, the wider the internal diameter of the gun. The bullets for shotguns range from single slugs with rifling grooves to numerous steel pellets. These pellets are close together at the muzzle but rapidly spread out as the distance from the gun increases. At close range shotguns are highly destructive, but they cause less injury as the range increases. Close range wounds should be inspected for embedded shotgun shell wadding or part of the shell, as these are potential sources of infection.

Metabolic response to trauma

Metabolism is the complex system of interrelated biochemical reactions and physiological responses required to maintain life. Traumatic injury is one of many stimuli that trigger a set of metabolic changes known as the stress response. The stress response is the body's damage control system, intended to maintain homeostasis on a cellular and systemic level, and to provide substrate for repair of injury. This response involves a set of hormonal and inflammatory signals that produces a hyperdynamic and hypermetabolic state, and mobilizes the immune response promoting haemostasis, wound healing and infection control. The systemic inflammatory response syndrome describes the pathological extension of the stress response.

Severity and type of injury determine the intensity and character of the stress response. The physiological response, described as the 'ebb and flow' phenomenon, is characterized by an early hypovolaemic hypoperfusion state with elevated oxygen consumption, and decreased urine output and blood pressures. During this time the inflammatory and hormonal mediators begin to accumulate. As the patient stabilizes a hyperdynamic state develops. This can take 12–24 hours and is characterized by increased cardiac output, oxygen delivery, tachycardia and fever with restoration of haemodynamic parameters. As oxygen delivery normalizes anaerobic metabolism diminishes, allowing clearance of lactate and base deficit. During this hyperdynamic state, carbohydrate, protein and fat metabolism are significantly altered. Hyperglycaemia, protein catabolism and lipolysis are prominent. The underlying purpose of these changes is the mobilization of energy and substrate for survival and wound healing. Glucagon, cortisol and catecholamines are the primary hormonal mediators of the stress response.

Fluid and electrolyte conservation

Decreases in fluid volume from haemorrhage lead to multiple signals directed at restoring circulating volume. Decreased atrial pressure releases atrial natriuretic peptide (ANP). Aortic and carotid body receptors trigger a neural response relating lower pressures. These inputs work through the hypothalamus and lead to release of antidiuretic hormone (ADH). The renin–angiotensin system is also activated and aldosterone is released. Aldosterone, ADH and ANP then work on the renal tubule to conserve sodium and water.

Hypermetabolism

The hyperdynamic state requires a great deal of energy; this results in an increase in basal metabolic rate (BMR). BMR can be predicted based on age, sex and body size. BMR can be measured using monitoring of respiratory gas exchange. Oxygen consumption in litres per unit time multiplied by 4.8 estimates the BMR in kilocalories per unit time. Metabolic rate is markedly increased in the postinjury patient. Severely injured patients may have a BMR that is 1.5 times normal.

Hyperglycaemia is common and is proportional to the type and severity of injury. In the ebb phase hyperglycaemia and hypoinsulinaemia are present. Later insulin levels rise but hyperglycaemia persists. Glucagon, epinephrine (adrenaline), norepinephrine (noradrenaline), vasopressin and angiotensin II are secreted in response to stress and promote glucose mobilization and production. These signals are also responsible for muscle catabolism. Usual controls of hyperglycaemia are lost and exogenous glucose and insulin no longer suppress hepatic gluconeogenesis. The stress response also increases peripheral insulin resistance, particularly in the skeletal muscle.

Following injury a catabolic state exists and muscle breakdown exceeds production. Nutrition with supplemental protein can help attenuate this response but will not prevent it. Skeletal muscle serves as the primary reserve pool of amino acids that are mobilized for gluconeogenesis in the liver. Glutamine is mobilized preferentially, and serves two unique purposes: it is utilized by gut enterocytes as an energy source and by the kidney for ammonia production. It is also used by fibroblasts and endothelial and immune cells.

The stress state is also lipolytic, and this state is not attenuated by exogenous nutritional sources, it is thought to be perpetuated by continued high sympathetic tone and catecholamine levels. Again, elevation in free fatty acids is proportional to injury severity. These high levels of fatty acids promote pathways of prostenoid production with resultant immunosuppressive effects. It is thought that limiting fat intake to 2 g/kg body weight per day may help to avoid this effect. In addition, modification of the specific fatty acid composition of dietary fat may be of value in modifying the immune response.

Oxygen and energy metabolism are also altered. Normally oxygen delivery exceeds consumption by a ratio of 5:1. This reserve is diminished in the stress state owing to decreased delivery from hypovolaemia, hypoxia and acute blood loss anaemia, and increased consumption from the hypermetabolic state. At the mitochondrial level there is uncoupling of oxidative phosphorylation, which can lead to a reliance on anaerobic pathways despite adequate oxygen delivery.

Because of the hypermetabolic state early nutritional support is essential to the care of the injured patient. Although catabolism is not completely eliminated, full nutritional support is essential for promotion of wound healing, prevention of infection, maintenance of lean body mass, and as a substrate for liver metabolism. Enteral nutrition is also essential to maintain gut mucosal integrity. Enteral nutrition is cheaper than parenteral nutrition, does not require central venous access, is physiological and serves as a trophic factor for enterocytes. There is a growing body of evidence to suggest that specific nutritional factors such as glutamine, arginine and omega-3 fatty acids are immunomodulators and that manipulating their concentrations in enteral feeds may allow the physician to alter the inflammatory response, immune system and nitrogen balance.

Mediators of the stress response to injury

Immune response after injury is complex and still poorly understood. Local tissue damage results in release of chemical signals that trigger local and systemic inflammation. This causes capillary leak, leucocyte infiltration, and local production of free radicals which can result in ongoing tissue damage. Systemically leucocyte demargination and mobilization from the bone marrow and spleen result in leucocytosis with a bias towards immature cell types, known as a 'left shift'. The leucocyte-produced cytokines involved in mediating inflammation include interleukin (IL) 1, IL-2, tumour necrosis factor (TNF) α and interferon (IFN) α.

TNF-α, or cachectin, so named because of its catabolic effects, can also be released by Kupffer cells. It is responsible for fever; anorexia; promotion of proteolysis, skeletal muscle wasting and negative nitrogen balance; tachypnoea; and tachycardia. TNF-α is also responsible for neutrophil release, margination and activation; as well as release of oxidants, superoxide, lysozyme and arachidonic acid metabolites. It is also responsible for increased vascular permeability, procoagulant effects and fibroblast proliferation needed for wound healing. At high levels it can also be associated with hypotension, organ failure and death.

IL-1 is released by macrophages and monocytes. It is similar to TNF-α and causes fever, is chemotactic for neutrophils, and increases endothelial permeability and catabolism. IL-1 stimulates release of IL-2 and IL-6 and shifts the liver to production of acute phase proteins. IL-2, produced by T-lymphocytes, stimulates cell-mediated immunity. It can be suppressed in trauma patients, and the degree of suppression has been correlated with mortality. In contrast, IL-6 is increased in trauma. It is released by fibroblasts and activated B-lymphocytes and promotes fibroblast, β-cell and hepatocyte proliferation.

IFN-α is produced by helper T-cells and macrophages. It is a priming agent for macrophage function, and increases production of other cytokines such as TNF-α, thus promoting the cascade. Other mediators include the leukotrienes, which are products of the lipoxygenase pathway or arachidonic acid metabolism. Produced in the vascular tissue and lung parenchyma by macrophages, mast cells and neutrophils, they serve as chemoattractants and smooth muscle vasoconstrictors. Prostenoids are also products of arachidonic acid metabolism and act as vasodilators, vasoconstrictors and platelet activators.

Specific organ system changes

Every organ system is affected by the stress state. The central nervous system is heavily involved in the response to injury; anxiety, pain and even ICU psychosis can lead to ongoing stimulation of the neurohumoral arm of the stress response. Cortical function may deteriorate as a result of regional ischaemia, accumulation of metabolic by-products or pharmacological agents. The lung is usually the first organ to fail. Carbon dioxide production is increased requiring increased minute ventilation; in addition, the lung is vulnerable to inflammatory effects of the postinjury state. Alveolar damage can trigger a self-perpetuating inflammatory process as the lung is rich in immune cells; this can result in the acute respiratory distress syndrome (ARDS). The liver undergoes major changes in response to injury as well. Hepatic blood flow is diminished; despite a diminished blood supply, hepatic activity is increased with gluconeogenesis, recycling of lactate through the Cori cycle, ureagenesis, and shifting of protein synthesis away from albumin towards acute phase proteins such as C-reactive protein, fibrinogen and caeruloplasmin. If these severe metabolic demands continue the liver becomes overwhelmed and cholestasis and steatosis can develop; if unchecked, frank liver necrosis may result. Gut function and mucosal integrity are diminished as blood flow is redirected away from the splanchnic system towards vital organs such as the heart and brain. In the early stages preservation of intravascular volume is key; oliguria results. As the stress response develops the kidney also needs to increase clearance of metabolic by-products to keep up with their increased production because of the body's hypermetabolic, catabolic and anaerobic state. Higher renal blood flow is required to meet these demands, and if not available because of hypovolaemia/hypotension acute tubular necrosis and azotaemia may result. In addition, circulating cytokines and inflammatory mediators as well as exogenous nephrotoxic agents may cause direct damage to renal tubular cells.

Modification of the stress response

Since cytokines are associated with many deleterious systemic effects, blockade of cytokine production or specific cytokine receptors has the potential to control the stress response. A number of cytokine blockade strategies have been attempted. Antibody to TNF-α has been used successfully to blunt the inflammatory response but has resulted in increased mortality in some trials. Other cytokine antibodies have been trialled alone or in combination with no clear success to date. Cyclo-oxygenase inhibition using non-steroidal anti-inflammatory compounds has the potential to inhibit the inflammatory cascade and has been met with some limited success. Early local and regional anaesthetic techniques have been found to limit the inflammatory response to surgical trauma and may be similarly beneficial in the injured trauma patient. Exogenous insulin and glucose control may be useful in attenuating catabolism; however, tight glucose control does not appear to be more beneficial than moderate control in trauma patients.

A recent area of intense research has been the use of beta-blockade in the injured patient. There is a growing body of evidence to suggest that administration of beta-blockers, particularly in the head-injured patient, may block the catecholamine surge associated with the stress response and decrease multisystem organ dysfunction and improve mortality.

Prehospital care and transport

Prehospital response

The prehospital response describes the systematic approach to the delivery of prehospital care to the injured patient. This system, referred to as emergency medical services (EMS), is usually organized on a local governmental level. The major goals of an EMS system are to provide rapid access by well-equipped vehicles, appropriate field management and rapid transport to an appropriate hospital. The EMS system is organized into three phases: the prospective phase, immediate phase and retrospective phase. The prospective phase encompasses funding, education of personnel, purchase of equipment and vehicles, and development and maintenance of treatment protocols. The immediate phase is initiated when the EMS system is activated, such as through a 999 call, and proceeds through assessment, triage, treatment and transport. Hospital personnel are notified of the situation and anticipated arrival using the MIVT (mechanism, injuries, vital signs, treatment) algorithm. The retrospective phase consists of review of every call after completion to ensure proper medical care was provided and for quality improvement.

Triage

Patients can be categorized into three groups based on their injuries: those that are rapidly fatal; those that are potentially fatal; and those that are not fatal. The first group constitutes approximately 5% of all patients, and are unlikely to benefit from prehospital care; the third group is the most common, representing 80% of all patients, and are unlikely to be harmed by delays in treatment. The second group, approximately 7–15% of patients, consists of those who have potentially fatal injuries. It is these patients who are most likely to benefit from diversion to a mature trauma centre and require most rapid transport.

Many triage protocols exist for identification of these patients. Such protocols aid in determining the optimal distribution of patients to hospitals in the area based on victim factors and the resources of the area such as number of, and distance to, specialty trauma centres. The American College of Surgeons Committee on Trauma (ACS-COT) has a protocol for this in *Resources for the Optimal Care of the Injured Patient*.

Outside of mass casualty and disaster management situations, categorization of traumas is generally based on patient acuity. Prior to the patient's arrival as much data from prehospital personnel should be gathered as possible. The components that determine level of acuity include clinical factors and mechanism of injury (MOI). Clinical criteria include physiological derangements of heart rate, blood pressure and level of consciousness. Clinical criteria have been repeatedly validated as useful triage tools. Scoring systems such as the Revised Trauma Score (RTS) (Table 14.2) are widely used in the field to quickly calculate the injury severity and predict the patient's medical needs. The RTS uses blood pressure, respiratory rate and the Glasgow Coma Scale (GCS) (Table 14.3) to calculate a coded

Table 14.2 Revised trauma score components

Glasgow Coma Scale	Systolic blood pressure (mmHg)	Respiratory rate	Coded value
13–15	>89	10–29	4
9–12	76–89	>29	3
6–8	50–75	6–9	2
4–5	1–49	1–5	1
3	0	0	0

Table 14.3 The Glasgow Coma Scale

Parameter	Score
Eye opening (E)	
Spontaneous	4
To speech	3
To pain	2
Nil	1
Best motor response (M)	
Obeys	6
Localizes	5
Withdraws (flexion)	4
Abnormal flexion	3
Extensor response	2
Nil	1
Verbal response (V)	
Orientated	5
Confused conversation	4
Inappropriate words	3
Incomprehensible sounds	2
Nil	1

Coma score: (responsiveness sum = 3–15 (E + M + V)).

value. MOI criteria in isolation have a very low predictive value in identifying severely injured patients. However, penetrating mechanisms such as stab wounds and gunshot wounds to the trunk or neck should be considered high-level traumas, as should blunt mechanisms that indicate significant force, such as falls over 4.5 metres, automobile versus pedestrian, and high-speed motor vehicle crashes. Other triage criteria such as age, associated medical conditions, environmental conditions and EMS personnel judgement should also be used during triage. Elderly patients in particular have multiple comorbidities, higher morbidity and mortality than their younger counterparts, sustain more severe injuries with minor mechanisms, and are therefore more likely to benefit from transfer to specialized trauma centres.

Because any triage system will not be perfect they should be evaluated routinely to ensure over- and undertriage rates are at acceptable levels. Undertriage occurs when a patient is not felt to have a serious injury but actually does; this can result in potentially preventable morbidity and mortality. Overtriage occurs when a patient is felt to have sustained a serious injury but actually has not; this can result in overutilization of finite resources. The ACS-COT has determined an acceptable undertriage rate to be ≤5%, while the acceptable overtriage rate may vary from 25% to 50%.

Methods of transport

Several options for transport exist including ground, helicopter or fixed wing flight, with the goal of transport to move patients quickly and safely. In the urban setting there appears to be no significant advantage for flight over ground transportation; rather, the main benefit of flight transport is for long-distance travel. However, air transport has unique complications relating to the patient; these include expansion of gases within body cavities such as pneumothoraces, and in potential spaces such as tissues with gas gangrene or penetrating eye injuries. Air can also expand within equipment such as air splints and endotracheal cuff balloons. Additionally, the space within most air transport vehicles is significantly smaller than that within ground transport vehicles.

Interhospital transfer

There are four ACS-designated trauma centre levels. Although all four levels are maximally committed to care of the trauma victim they are distinguished from each other by available resources and personnel capable of caring for the acutely injured. It is imperative for the initial treating physician to recognize that some trauma victims may require diagnostic or therapeutic modalities beyond the scope of the receiving hospital. The need for transfer is based on patient status, injuries and comorbidities. Injuries that will require prompt treatment by surgical subspecialty services are best treated at designated trauma centres. Examples include damage to the central nervous system; major chest injuries; severe abdominal or pelvic injuries; complex or open fractures; or patients requiring revascularization. Multitrauma patients and those at the extremes of age may also benefit from transfer.

Transfer agreements with designated trauma centres should be present in all hospitals with limited resources for injured patients. Transfer agreements are established protocols between hospitals that ensure rapid and efficient passage of pertinent patient information prior to actual patient transport. This should include patient identification, history and physical examination findings, diagnostic and therapeutic procedures performed and their results. The accepting trauma physician may also suggest possible diagnostic or therapeutic manoeuvres prior to transfer such as intubation or insertion of a nasogastric tube, Foley catheter or thoracostomy tube. It also allows for mobilization of resources such as operating theatre or ICU beds at the receiving hospital. The physicians involved should also discuss the mode of transportation, accompanying personnel and equipment that may be needed for optimal transfer. Derangements in the patient's physiological status should be identified and stabilized as best as possible prior to transfer.

Transfers do have the potential for instability and these risks must be balanced against the benefits of a higher level of care. Risks can be minimized with the use of proper equipment, personnel and planning. In addition to the medical aspects of interhospital transfers, physicians must comply with certain federal and local legal regulations.

Assessment of the injured patient

Prehospital phase

Care of the injured patient begins in the prehospital setting with a tightly integrated emergency medical response system designed to provide immediate access to life-saving medical care. In most locations, this involves a first response team of paramedics who provide basic life-support measures, external compression of bleeding, splinting and spine stabilization within minutes of a traumatic event. Often, these prehospital personnel can also perform advanced procedures, including intravenous access, intubation and pleural decompression. The patient is stabilized and rapidly transported to a trauma centre, where resuscitation is continued and injuries are identified and treated expeditiously.

Hospital phase

The transfer of patient care from prehospital to in-hospital teams should occur in an organized fashion with a report that rapidly details the mechanism of injury, obvious injuries, the patient's vital signs, and treatments provided by prehospital providers (MIVT). The report becomes the formal mechanism by which transfer of care occurs. Assessment and treatment then follow a logical sequence based on clinical judgement and the principles of the Advance Trauma Life Support (ATLS) programme developed by the ACS-COT. A team approach involving physicians, nurses, technicians and trained paramedical personnel allows resuscitation to occur while ongoing evaluation is conducted.

The team must have a captain responsible for organizing the team efforts and performing the primary survey. He or she should rapidly determine the nature of diagnostic tests or interventions required by the patient, and the order in which those activities should be performed. Various tasks can be accomplished simultaneously by different team members, all of which should be co-ordinated by the team leader.

Preparation for patient arrival

Prior to patient arrival it is essential that monitoring devices, warmed intravenous fluids, blood products, multiple intravenous access devices, and Vacutainers for blood samples should be immediately available. Adjunctive devices – suction canisters, intubation and ventilation equipment, urinary and nasogastric catheters, splinting or casting material, pelvic binders and surgical instrument trays – should also be easily accessible.

Personal protective gear (PPG) should be donned by all members of the treatment team. Precautions should include occlusive gown, gloves, face shield or at a minimum eye protection, a mask and surgical cap if any sterile procedures are anticipated, and shoe covers. PPG is worn not only for the protection of the patient but also for the protection of the staff from blood-borne infection, which ranged from 1.4% to 1.9% for HIV and from 3% to 8% for hepatitis C in recent studies of trauma patients.

Primary survey

The primary survey consists of the ABCDEs (airway, breathing, circulation, disability and exposure/environment). This ABCDE sequence has been derived to address life-threatening injuries most rapidly; as such, airway obstruction, which causes death in 1–5 minutes, is first, while less rapidly lethal problems such as hypoxia and hypotension are second and third. Upon arrival patients are visually surveyed for injuries; devices such as non-invasive blood pressure cuffs, pulse oximetry and electrical cardiac monitors are used to obtain vital signs; large-bore intravenous access is established; and the patient is completely disrobed. Adjuncts to the primary survey include plain radiographs of the chest and pelvis, and focused assessment with sonography for trauma (FAST), including views of the abdomen and pericardium.

During evaluation of the ABCDEs full spinal precautions should be maintained, and resuscitation should be ongoing. A type and cross-match should be the first priority in obtaining blood samples. Blood should also be drawn to obtain haematological, chemical and coagulation profiles, and arterial co-oximetry. Pregnancy tests should be done in all females of childbearing age.

Airway

The clinician must assess whether the airway is patent, whether it is protected and whether there is an evolving injury that may have an impact on future patency of the airway. The patient who responds when questioned has an adequate, protected airway. The patient who is unconscious or who demonstrates respiratory insufficiency is at risk for aspiration, hypoxia and hypercarbia, and requires definitive airway control.

The evaluator should perform simple manoeuvres such as the jaw thrust or chin lift, and suctioning of the mouth and oropharynx to clear the airway, and high-flow supplemental oxygen should be provided. Supplements such as the nasal trumpet and oral airway can be considered. During this time, any interventions require an awareness of the possibility of cervical spine injury.

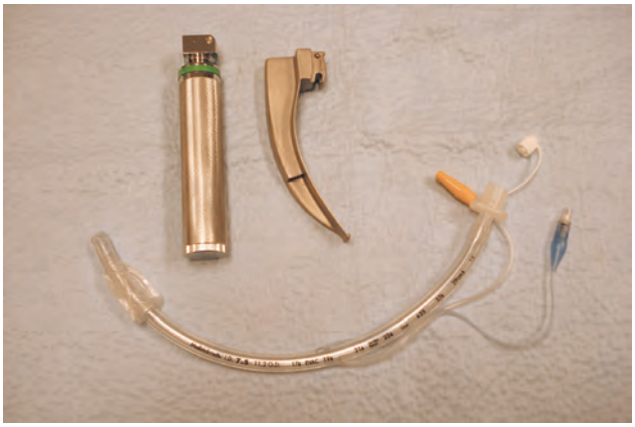


Figure 14.2 Airway tools: laryngoscope and endotracheal tube.

If there continues to be evidence of respiratory insufficiency after these manoeuvres, or if the patient remains unable to protect the airway, definitive control with endotracheal intubation or cricothyroidotomy is necessary (Figure 14.2). Intubation in the comatose patient can generally be accomplished without sedating or paralysing medications. Increased muscular tone and biting make sedating and paralytic agents necessary, as does intubation of the conscious patient. In these cases rapid sequence intubation (RSI) is utilized. RSI is a technique of administering boluses of a sedative with/without a paralytic agent while using cricoid pressure to prevent aspiration with little or no preoxygenation, again to prevent gastric insufflation and aspiration. It is important to obtain a very quick neurological assessment prior to intubation: if 5–10 seconds can be spared, an assistant should ask the awake patient to move his or her limbs, briefly assess sensation and perform a pupillary examination. In the comatose patient response to painful stimuli should be ascertained and a rectal and pupillary examination should be performed. Once intubation is accomplished, appropriate positioning of the endotracheal tube (ETT) should be confirmed clinically, with a carbon dioxide detector, and radiographically. Care should be taken to rule out right mainstem intubation.

In some cases traditional orotracheal intubation can be difficult or impossible; in these cases conversion to a surgical airway should occur immediately. Cricothyroidotomy is the preferred surgical airway in adults. The patient's neck is exposed while maintaining strict cervical spine immobilization, the cricothyroid membrane is palpated between the thyroid and cricoid cartilages, a vertical midline incision is made in the skin, the membrane is exposed and divided in the transverse direction, and the ETT or tracheostomy tube is placed into the airway.

In certain patients the airway may be patent upon initial inspection but may be in danger of becoming compromised. These high-risk patients and those who require prolonged transport may benefit from elective placement of a definitive airway. Finally, agitated and combative patients or those with acute alcohol or drug intoxication may be unable to protect their airway and should be promptly intubated as well.

Breathing

After establishing a controlled airway it is necessary to ensure adequate gas exchange. Life-threatening conditions that should be diagnosed on primary survey include tension pneumothorax (Figure 14.3), open pneumothorax or sucking chest wound (Figure 14.4), massive haemothorax (Figure 14.5) and flail chest (Figure 14.6). These should be identified by absence of breath sounds on clinical examination in most cases and treated with tube thoracostomy, pain control and mechanical ventilation.

Once the most immediately life-threatening injuries have been addressed, adjuncts to the breathing assessment such as arterial blood gas and plain radiographs can be obtained. Arterial blood gas is especially important in the head-injured and comatose patient as significant hypercapnia may be present and is difficult or impossible to assess clinically. Chest radiograph (CXR) is an essential adjunct to the primary survey and can aid in diagnosis and treatment of both airway and breathing problems. It can be used to assess correct positioning of definitive airways and chest tubes and to detect simple pneumothorax, haemothorax, pulmonary contusion and broken ribs. CXR can also reveal a widened mediastinum, which may indicate thoracic aortic injury; sternal or thoracic spinal fracture; and an enlarged cardiac silhouette, which might indicate the presence of pericardial fluid.

Circulation

After airway and breathing complications are addressed, circulation is the next priority. Blood pressure, pulse, skin perfusion, urine output, mental status and central venous pressure are clinically useful indicators of haemodynamic condition. The goal

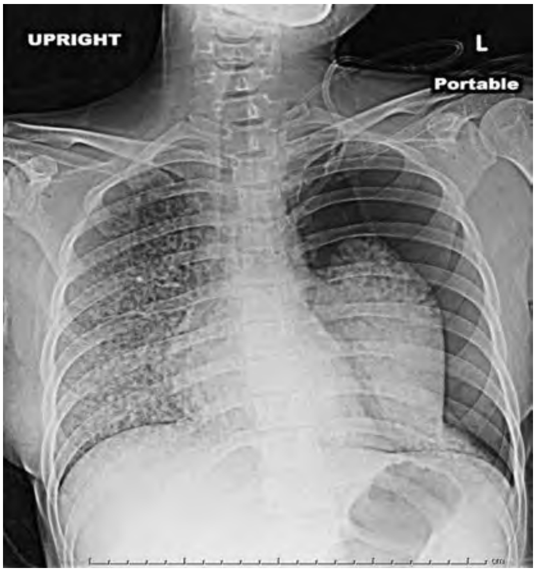


Figure 14.3 Chest radiograph demonstrating tension pneumothorax; note the deviation of the trachea and heart away from the pneumothorax to the contralateral side.

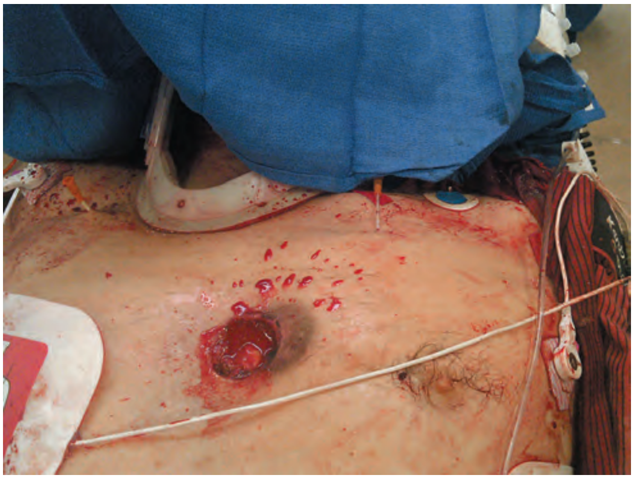


Figure 14.4 Sucking chest wound from close range shotgun blast.

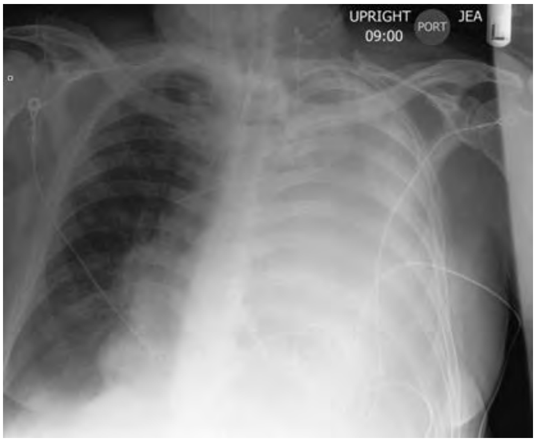


Figure 14.5 Chest radiograph demonstrating massive haemothorax of the left chest; note the deviation of the mediastinal structures to the contralateral side.

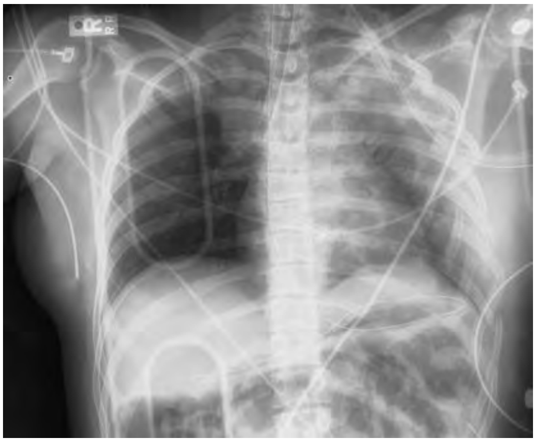


Figure 14.6 Chest radiograph demonstrating flail segment of the left lower ribs with associated haemothorax causing increased opacification of the left hemithorax.

is to determine whether the patient is in shock. It should be kept in mind that elderly patients often lack the intrinsic ability to mount a sympathetic response to haemorrhage or may be on beta-blockers, preventing normal tachycardia, and be in a state of occult hypoperfusion with apparently normal vital signs.

Types of shock include cardiogenic, redistributive, obstructive and hypovolaemic. Hypovolaemic shock due to blood loss is the most common cause and should be assumed to be the source of hypotension in all trauma patients. However, both types of obstructive shock should be rapidly ruled out in all hypotensive trauma patients. These are tension and cardiac tamponade. Cardiac tamponade results from external compression of the heart by blood or fluid resulting in hypotension. Clinical indications of tamponade include restlessness, inability to lie supine, distended neck veins, muffled heart sounds and pulsus paradoxicus. Diagnosis can be confirmed with pericardial ultrasound, which is highly sensitive and specific, nearing 100% in most studies. If tamponade is confirmed, treatment is emergency surgery. If no surgeon is present pericardiocentesis can be used to temporize the patient until transfer to a surgeon can occur.

Haemorrhagic shock is classified according to severity, amount of blood lost and extent of organ failure (Table 14.4). Areas where significant blood loss can occur include external loss and bleeding into the chest, the abdomen, the pelvis or retroperitoneum, and the extremities. Ongoing fluid requirements or lack of response to resuscitation indicates significant haemorrhage and necessitates immediate measures to identify and control the source of bleeding. Sources of external bleeding should be controlled with direct pressure or tourniquets, and fractures should be splinted. Bleeding into the chest is diagnosed primarily with CXR or after chest tube placement. Pleural effusion, appearance of elevated hemidiaphragm, and white-out of the affected lung field are all indications of haemothorax and should be treated with immediate placement of a thoracostomy tube. Evidence of significant abdominal trauma on physical examination includes seat belt signs, abdominal wall haematomas or lacerations, or abdominal distension. Haemoperitoneum may also cause peritonitis and significant abdominal pain. Patients with peritonitis should be taken urgently for laparotomy regardless of their haemodynamic status. Patients without peritonitis should undergo FAST. FAST consists of four ultrasonographic views: the pericardium (subxiphoid or parasternal), hepatorenal fossa, splenorenal fossa and around the bladder. FAST is easily accessible, can be performed in less than 2 minutes and is portable, repeatable, inexpensive and non-invasive. A positive FAST (Figure 14.7) in a hypotensive patient is an indication for immediate laparotomy. However, an equivocal or negative FAST, especially in the hypotensive patient, cannot reliably rule out intra-abdominal bleeding. In the unstable patient, if the FAST is negative or equivocal, diagnostic peritoneal aspiration (DPA) should be performed. Aspiration of 10 mL of blood or more is considered positive. If no blood returns consideration can be given to diagnostic peritoneal lavage (DPL). If gross blood or succus is noted, this is a positive finding if at least 75% of the infusate is returned. If >100 000 red blood cells (RBCs)/mm^3 (blunt) or 1000 RBCs/mm^3 (penetrating), >500 white blood cells (WBCs)/mm^3 or food particles are found this is also considered a positive finding necessitating laparotomy. If FAST and DPA are negative in patients with abdominal trauma, retroperitoneal and pelvic bleeding are potential sources of haemorrhage. Clinical examination includes palpation of the pelvis for lateral and anteroposterior stability. A stable pelvis on examination with full active range of motion of both hips nearly excludes significant pelvic fracture. In comatose patients this may not be possible, and the mainstay of diagnosis is plain film of the pelvis (PXR). If pelvic fracture is diagnosed on clinical examination or PXR and hypotension is present the pelvic ring should be reapproximated using an external compression device; these include pelvic binders, external fixators and sheets. The placement of stabilization devices closes the pelvic ring, decreases pelvic volume and tamponades bleeding.

Table 14.4 Classification of shock in adults based on initial presentation (for a 70 kg man)

	Class I	Class II	Class III	Class IV
Blood loss (mL)	≤750	750–1500	1500–2000	≥2000
Blood loss (% blood volume)	≤15	15–30	30–40	≥40
Pulse rate	<100	>100	>120	>140
Blood pressure	Normal	Normal	↓	↓
Pulse pressure	Normal or ↑	↓	↓	↓
Capillary refill	Normal	Delayed	Delayed	Delayed

Improvements in prehospital care have resulted in more frequent transport of patients *in extremis* to trauma centres. Rapid assessment of these patients is required to decide who will benefit from aggressive resuscitative efforts. Resuscitative thoracotomy or emergency department thoracotomy (EDT) is performed to address cardiovascular collapse. It allows evacuation of haemothorax/pneumothorax, release of tamponade, repair of cardiac and intrathoracic vessel injuries, evacuation of air embolus, clamping of hilar injuries, cross-clamping of the aorta, and performance of open cardiac massage (Figure 14.8). In general, EDT is indicated for patients with witnessed arrest with cardiopulmonary resuscitation (CPR) <5 minutes

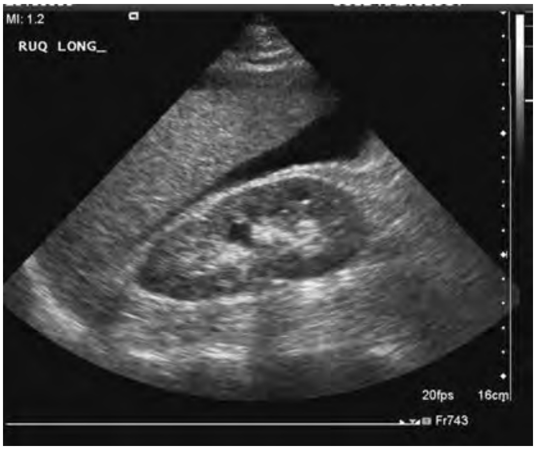


Figure 14.7 Positive focused assessment with sonography for trauma: the dark strip represents free fluid in Morrison's pouch between the liver and the right kidney.

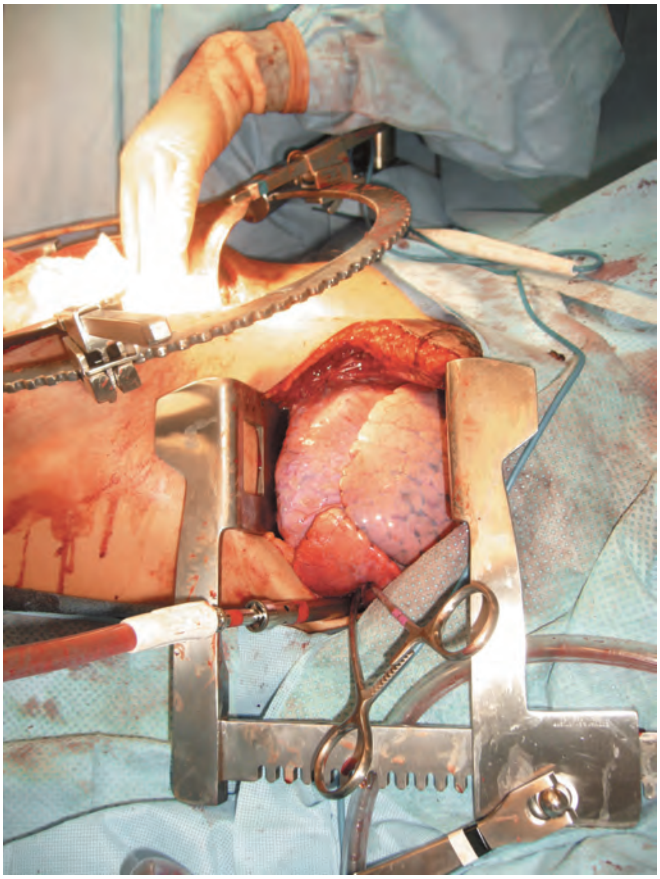


Figure 14.8 Resuscitative thoracotomy with aortic cross-clamp in place.

following penetrating thoracic trauma, witnessed arrest upon or after arrival following blunt trauma if signs of life were present in the field, and persistent hypotension due to tamponade or intrathoracic haemorrhage. Outcomes are best with penetrating cardiac injury, followed by penetrating non-cardiac thoracic trauma. Outcomes following penetrating and blunt abdominal trauma are poor. Outcomes in children are similar to adults, and indications for paediatric EDT are similar to those for adults.

Resuscitation strategies

Treatment of shock begins with immediate insertion of two large-bore intravenous lines, ideally placed during the prehospital phase; however, placement of such lines should not delay transport. If peripheral access is impossible an intraosseous or central line can be inserted. Intraosseous lines can be placed into the tibia, humerus, sternum or clavicle. The rate of infusion and the range of medications that can be infused through an intraosseous line are excellent and comparable to peripheral venous lines. As with any invasive procedure, complications from intraosseous insertion do occur, including compartment syndrome, osteomyelitis and fracture; however, complication rates are very low. Central venous catheters are also an option. Locations for central lines include jugular, subclavian and femoral. Care should be taken to avoid placement of central lines when significant vascular injury proximal to the access site is suspected. Other choices for venous access include saphenous vein cutdown at the ankle or groin, or cutdowns to the cephalic or basilic vein in the antecubital fossa.

Resuscitation should begin with infusion of 2 litres of warmed crystalloid if hypotension is present. If the systolic blood pressure fails to improve, O-negative packed RBCs (PRBCs) should be transfused. However, in selected patients with penetrating torso trauma a strategy of permissive hypotension may be beneficial. A study in which patients were randomized to traditional or delayed resuscitation revealed a significant survival benefit in patients with delayed fluid resuscitation, as well as fewer complications and shorter hospital lengths of stay. However, few studies have been able to replicate the positive effects of delayed resuscitation. At this time, there is still active debate as to the benefit of permissive hypotension and the mainstay of treatment of haemorrhagic shock remains active fluid resuscitation and immediate source control.

When transfusion of large volumes of blood is necessary the optimum ratio of PRBCs to fresh frozen plasma and platelets is also an area of active research. Massive transfusion is variably defined, but is generally accepted to be any patient requiring >10 units of PRBCs within the first 24 hours of hospitalization. In these cases, dilution of fibrinogen, platelets and clotting factors occurs as whole blood is lost, and is replaced with only PRBCs and crystalloid. Both military and civilian studies reveal decreased mortality in a step-wise fashion with increasing plasma to red cell ratio, with optimal results approaching a ratio of 1:1. Protocolizing transfusion and transfusion ratios has increased in popularity recently. The goal of massive transfusion protocols (MTPs) is to standardize the replacement of platelets and clotting factors in the optimum ratio to red cells to prevent coagulopathy, but avoid increased risks of multisystem organ dysfunction and increase efficiency of transfusion. MTPs are helpful in reducing time to transfusion, which results in faster blood/component replacement and may decrease crystalloid volumes. Reduced times to transfusion of first blood products exclusive of the transfusion ratio has also been associated with decreased mortality.

An adjunct to a balanced transfusion strategy is the use of pharmacological agents to treat coagulopathy, including recombinant human factor VIIa and prothrombin complex, which contain factors II, VII, IX and X. There is some evidence that use of these agents may result in decreased transfusion requirements and lower rates of transfusion-related organ failure among certain trauma patients.

Disability

After securing the airway and restoring oxygenation, ventilation and adequate circulation, the patient's neurological status should be ascertained. This consists of the GCS score (Table 14.3), pupillary examination and gross determination of movement and sensation. The neurological examination should be deferred until restoration of adequate oxygenation, ventilation and circulation as any derangements can render the examination inaccurate. Other factors that may alter the GCS score and pupillary examination include intoxication with drugs or alcohol, hypothermia and hypoglycaemia. After restoration of oxygenation and circulation, any pupillary abnormalities or alterations in GCS score should be presumed to be the result of traumatic brain injury until advanced imaging has definitively ruled it out. Patients with a GCS score of ≤8 or lateralizing signs are very likely to have an intracranial mass lesion requiring prompt surgical intervention.

Any patients with presumed brain injury should undergo CT scan of the head as soon as possible after the primary survey is completed. Patients with a GCS score of ≤8 and any two of age >40, hypotension or lateralizing signs are at significantly increased risk of intracranial hypertension, and should undergo placement of an intracranial pressure-monitoring device regardless of CT findings. In patients with presumed or confirmed brain injury neuroprotective resuscitation should begin as soon as possible. This includes mild hyperventilation to a $P\text{aCO}_2$ of 32–35 mmHg, circulating blood volume, avoidance of hypotonic solutions and maintenance of adequate oxygenation.

Full spinal precautions must be maintained with protection and immobilization of the entire spine throughout the evaluation process. Previously in patients with definite evidence of blunt spinal cord injury a steroid bolus followed by continuous infusion for 24–72 hours was recommended. However, a growing concern over limitations of the studies these guidelines were based on and failure of subsequent studies to demonstrate a definite benefit to steroid administration have led to the removal of the recommendations for the use of steroids in spinal cord injury from the 8th edition of the ATLS manual.

Exposure

In order to complete the primary survey the patient must be fully disrobed to reveal all injuries, and log-rolled maintaining full spinal precautions to assess the back and perineum. After complete visual inspection, the patient should be covered as quickly as possible with warm blankets to prevent hypothermia.

Interestingly, several studies of victims of non-traumatic cardiac arrest and patients undergoing cardiac or neurosurgical procedures have shown significant improvements in outcome with induction of mild to moderate hypothermia with core temperatures of 32–34°C. However, results in trauma patients have shown a significantly higher rate of multisystem organ dysfunction and mortality. Traumatic brain injury, in particular, has shown no benefit in survival or neurological outcome with therapeutic hypothermia. At this time hypothermia should still be considered detrimental in both the adult and paediatric trauma population.

Secondary survey

The secondary survey is a quick but comprehensive head-to-toe physical examination. The secondary survey may include simple or advanced imaging such as extremity plain films or CT scans. The secondary survey also includes a brief directed history, which should follow the AMPLE model of allergies, medications, past illnesses/operations, last meal, and events/environment related to the injury. The history should be directed at detection of pre-existing conditions or medications that may alter management of life-threatening injuries. The history should also include information from the prehospital providers regarding the mechanism of injury, which can help the physician to judge those areas of the body that may have sustained the greatest energy transfer so that subsequent investigations can be directed appropriately.

The head and neck examination should include inspection and palpation to detect lacerations, haematomas, fractures and tenderness, otoscopic evaluation to detect haemotympanum, and assessment of all cranial nerves. Any lacerations should be rapidly closed with sutures or staples to prevent ongoing blood loss. The primary adjuncts to the head and neck assessment include radiographs of the cervical spine and CT scans of the head, face and neck. CT of the neck can also be performed with intravenous contrast to assess the soft tissues and carotid and vertebral arteries with high sensitivity.

The majority of the thoracic examination is undertaken in the primary survey, when physical examination and CXR will detect most injuries. The two major exceptions are thoracic aortic injury and thoracic spine fracture. Both may present with widened mediastinum and back pain. Patients with aortic injury may also have loss of aortic contour, apical capping, tracheal or nasogastric tube deviation, and loss of the aortopulmonary window on CXR. Any patient with significant external evidence of thoracic trauma should undergo electrocardiogram (ECG) and CT scan of the chest with intravenous contrast. Patients with ECG changes may have blunt cardiac injury and should be monitored with telemetry for at least 24 hours. The CT scan of the chest will reliably diagnose spinal fracture, give additional information on parenchymal lung injury (contusion, laceration, occult pneumothorax), and is now the screening examination of choice for thoracic aortic injury.

The abdomen and pelvis should be inspected and palpated during the secondary survey. A rectal examination should be performed. If the prostate is noted to be high riding, in the presence of urethral bleeding or perineal/scrotal haematoma, a retrograde urethrogram should be performed to rule out urethral injury prior to Foley catheter placement. If there is evidence of injury, a suprapubic catheter can be placed. As stated previously, unstable patients may undergo DPA/DPL as an adjunct to the primary or secondary survey. In stable patients an adjunct to the secondary survey is a CT scan of the abdomen and pelvis. CT scan of the abdomen and pelvis with intravenous contrast can detect fractures of the thoracolumbar spine and pelvis, free fluid or air in the peritoneal cavity, and solid organ injury. Additionally, in the presence of concerning pelvic fractures or gross haematuria, delayed phase images can be obtained to examine the urinary collecting system. CT scan of the abdomen and pelvis cannot be relied upon to detect hollow viscous injury. Some signs of hollow viscous injury on CT scan include free fluid without solid organ injury, focally thickened bowel wall and free intraperitoneal gas; however, these signs are neither sensitive nor specific. The gold standard for detection of hollow viscous injury remains laparotomy or serial observation.

The patient should also be log-rolled while maintaining full spinal precautions to inspect the back, spine and posterior head and neck for wounds, crepitus and any tenderness to palpation.

A complete musculoskeletal survey should also be performed during the secondary survey; all joints should be assessed for passive and active range of motion, long bones should be assessed for instability, the limbs should be examined for external signs of trauma or violation of the joints, and a sensory examination of all dermatomes should be performed to assess for peripheral nerve injury. A complete assessment of all peripheral pulses should also be performed. If any asymmetry is noted, and in cases of obvious fracture or deformity even if pulses are present in the affected limb, an ankle–brachial index (ABI) or brachial–brachial index (BBI) should be performed after reduction of fractures to rule out occult vascular injury. The blood pressure should be taken below the level of injury and compared with the systolic blood pressure in the unaffected upper extremity. If the ABI or BBI is >0.9 occult arterial injury is unlikely; if it is <0.9 further imaging should be obtained. Ultrasound and CT angiography are commonly used modalities. Extremities must be monitored frequently for the development of compartment syndrome. Compartment pressures should be measured if any symptoms such as pain out of proportion to examination or with passive range of motion or signs such as oedema or tense compartments develop. Elevated compartment pressures are an indication for emergency fasciotomy.

Once the secondary survey has been completed, if the decision to transfer the patient to another facility has been made, transfer should not be delayed for adjunctive imaging, and an appropriate level of transport care should be determined and arranged.

Specific injuries

Soft-tissue injuries of the neck

The neck is an anatomically complex region that acts as a conduit for multiple structures, including vital components of the vascular, neurological, respiratory and digestive systems. The neck is enveloped by the platysma and the superficial fascia, and by the deep cervical fascia, which includes the pretracheal and prevertebral planes. The space between the pretracheal layer and the prevertebral layer is called the visceral compartment, where all the vital structures are located. Clinically, the neck is divided into three surgical zones (Figure 14.9). Zone 1 is bounded by the clavicles and cricoid cartilage; zone 2 lies between the cricoid cartilage and the angle of the mandible; finally, zone 3 extends from the angle of the mandible to the base of the skull. Zone 1 and 3 injuries are difficult to expose surgically and are typically investigated with imaging and panendoscopy. All zone 2 injuries historically underwent mandatory operative exploration. More recent experience has shown good outcomes with a policy of selective operative management. Currently, the primary determinant of operative versus conservative management is clinical presentation. Regardless of location of injury, patients with hard signs of vascular or aerodigestive injury require surgical exploration; those with soft signs should undergo imaging; those with no signs of injury may be observed or discharged.

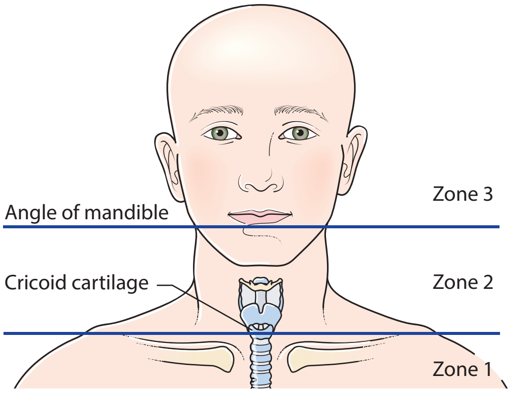


Figure 14.9 Classification of the zones of the neck.

Initial evaluation

The propensity of respiratory compromise in patients with vascular or tracheal injuries makes the early establishment of a secure airway a priority. Approximately 10% of patients with neck injuries will have some degree of airway compromise. Haematoma, laryngeal fractures, soft-tissue swelling or intraluminal bleeding may cause upper airway obstruction. Many patients with blunt injuries will have major associated craniofacial trauma, requiring intubation for airway management. Endotracheal intubation is mandatory in patients with significant obstructive signs and should be considered as a prophylactic measure when injuries are deemed capable of producing obstruction. Direct laryngoscopy may not be tolerated well in the awake patient, and medications used in rapid sequence intubation may convert an urgent airway into an emergency airway. Direct laryngoscopy may also be technically challenging owing to distortion of the airway from overlying soft-tissue damage. Because of this, the operating theatre has several advantages over the emergency department for airway management in patients with neck injury. It is a controlled environment, may have more access to advanced airway equipment, and conversion to a surgical airway can occur with the aid of extra personnel, equipment, better positioning, improved lighting and sterile technique. Despite meticulous planning, the establishment of a surgical airway is often necessary in patients with significant neck injury. Cricothyroidotomy is a good option for the unstable patient, but should only be used if endotracheal intubation fails in the stable patient. Additionally, preservation of as much intact airway as possible is important if there is a tracheal injury.

Patients in hypovolaemic shock should be resuscitated and obvious external haemorrhage from cervical wounds should be controlled by direct external pressure. Overlooked occult haemorrhage from seemingly innocuous penetrating neck wounds is a dangerous pitfall. Wounds at the base of the neck may produce massive internal bleeding into the mediastinum or pleural space, or into an adjacent tracheal wound. Massive haemorrhage from these injuries can occasionally be controlled by direct pressure, or by insertion of Foley catheters in the wound tract to tamponade bleeding (Figure 14.10), while arranging

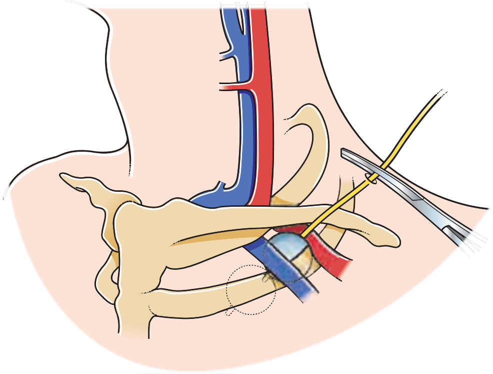


Figure 14.10 Balloon tamponade of a zone 1 neck injury with Foley catheter. (From Asensio JA, Demetriades D (eds) *Techniques in Complex Trauma Surgery*. Philadelphia, PA: W.B. Saunders, 2003.)

transport to the operating theatre. Immediate anterolateral thoracotomy or sternotomy for proximal vascular control should then be performed for vascular control and repair.

After securing the airway and ensuring ventilation and circulation are intact all patients should be assessed for hard and soft signs of injury. All patients with neck injuries should be visually inspected, including the oropharynx, and the neck should be auscultated for bruits and palpated for thrill or crepitus. All patients should also undergo a careful neurological evaluation, including cranial nerves, particularly VII and IX–XII, and motor and sensory examinations. All findings should be clearly documented, particularly if operative management is likely. Hard signs of injury include expanding or pulsatile haematoma, bruit/thrill, active bleeding, hypotension, air bubbling through the wound, stridor, unexplained focal neurological deficits and airway compromise (Table 14.5). Soft signs include small volume haemoptysis or haematemesis, dysphagia, dysphonia, subcutaneous emphysema and isolated nerve injury. Unstable patients and those with hard signs of injury should be taken for operative exploration. For clinically stable patients with soft signs of injury, invasive and non-invasive studies (radiographs, contrast radiographs, arteriography, CT, endoscopy) may serve several purposes. They may be used for localization of a known or suspected lesion and aid in planning the operative approach. They may also be used to exclude potential injuries. In the case of arteriographic embolization, the procedure may also be therapeutic.

Table 14.5 Indications for operative intervention in the management of neck injuries

Type of injury	Indication
Vascular	Hypotension/shock Active external bleeding Large or expanding haematoma Lack of distal pulse Bruit/thrill Significant haemothorax or haemomediastinum
Aerodigestive	Massive subcutaneous emphysema Air bubbling from the wound Large haemoptysis or haematemesis Stridor
Neurological	Hemispherical deficit or hypoperfusion

Chest films are routinely obtained in all trauma patients. Occult haemorrhage into the chest, associated pneumothorax or cervical soft-tissue air can be seen on radiographs. Neck films may also be useful for detecting foreign bodies (e.g. glass) and in the assessment of gunshot trajectories. Arteriography can be performed when vascular injury is suspected and has been found to be very accurate. The reliability of arteriography in excluding mediastinal vascular injuries has been questioned, however, owing to the size and orientation of the aorta and great vessels, and arteriography is ill suited for detection of venous injuries. CT angiography (CT-A) of the neck is fast becoming the initial diagnostic examination of choice. CT-A gives information on both arterial and venous vasculature, bony structures, soft tissue and most importantly the tract of injury in penetrating trauma. If the tract of injury is far from vital structures, work-up can be considered complete. CT-A of the neck is diagnostic for vascular injuries with sensitivities approaching 100%. It is less accurate for evaluation of the aerodigestive tract; however, if the tract is suspicious, oesophagoscopy, bronchoscopy and contrast oesophagography can be performed. In patients with wounds superficial to the platysma or with complete absence of any clinical signs no further evaluation is generally required.

The use of contrast oesophagography in the evaluation of cervical oesophageal injuries is controversial, as the sensitivity is not as high as it is for the intrathoracic oesophagus in which effluent from a laceration or perforation tends to be less contained. Several investigators have found the incidence of missed injuries associated with this method to be as high as 50%. Sensitivity can be increased by distending the oesophagus with water-soluble contrast under slight pressure through a proximally placed nasogastric tube. Similarly, oesophagoscopy has not proven to be a reliable means of excluding injury, with a reported sensitivity in several small studies of only 40–50%. Using a combination of oesophagoscopy and oesophagography however, may increase sensitivity to 75–85%.

Penetrating neck injuries

Zone 1 injuries occur at the base of the neck and do not allow easy accessibility from an operative standpoint. Injuries in zone 1 may involve proximal carotid, subclavian or innominate vessels, and patients are at risk for exsanguinating haemorrhage, which may be occult if the blood tracks into the chest or mediastinum. Asymptomatic patients can be evaluated by CT-A or angiography. Most studies suggest that asymptomatic zone 1 patients with negative imaging may be safely managed non-operatively. However, as many as 39% of patients with zone 1 vascular injuries present with unexplained shock. The management strategy for symptomatic or unstable zone 1 injuries consists of early operative exploration. A variety of incisions have been used to gain exposure and vascular control for thoracic inlet injuries (Figure 14.11a,b). The operative approach will be determined by the specificity of localizing signs, preoperative arteriography or associated intrathoracic injuries. Transverse clavicular incisions are used for access to distal subclavian injuries. Injuries to the mid-portion of the subclavian vessels

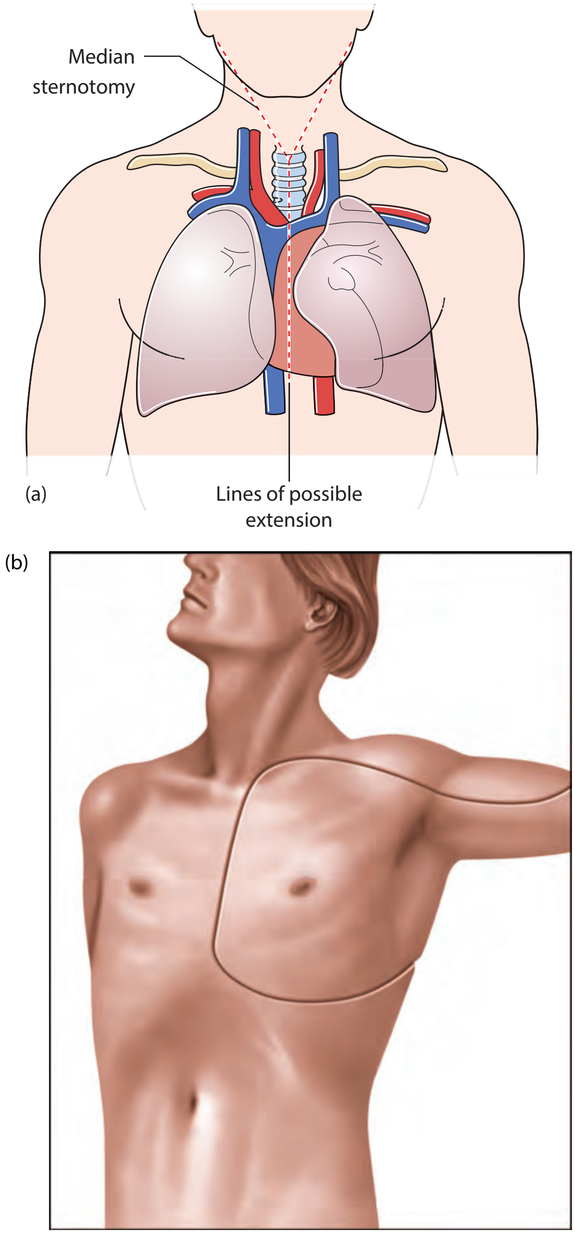


Figure 14.11 Operative approaches for injuries to the neck. (a) Anterior sternocleidomastoid incision with optional mediastinal extension excellent for zone 1 and 2 injuries. (b) Trapdoor with modification to include extension to left thoracotomy, another option for zone 1 injuries particularly the left subclavian artery and vein. (From Hoyt *et al.*, *Surg Clin North Am* 2001;**81**:1299–330.)

will usually require either a clavicular osteotomy or resection of the clavicle to obtain adequate exposure. Proximal access to right subclavian or innominate vessels is best obtained through a median sternotomy with supraclavicular lateral extension for additional exposure. Left proximal subclavian injuries will generally require a posterolateral thoracotomy for control. 'Trap door' or 'book' incisions combine an anterior thoracotomy with a median sternotomy and supraclavicular extension for wide exposure to left-sided and aortic arch injuries. More distal carotid injuries can be best approached via an anterior sternocleidomastoid incision.

It is injuries to zone 2 that have generated the greatest controversy regarding their management. Exploration of the vital structures in this location is relatively simple surgically and can generally be accomplished via a standard anterior sternocleidomastoid incision. The historical policy of mandatory exploration of all injuries in zone 2 that penetrated the platysma resulted in a significant number of negative explorations, and careful retrospective review of patients with positive and negative explorations found that physical examination findings were quite sensitive in detecting patients likely to benefit from exploration. Current recommendations are for immediate exploration for all patients with hard signs of injury (Table 14.5); CT-A with adjunctive angiography, oesophagography, bronchoscopy and oesophagoscopy on a case-by-case basis based on location and tract of injury on CT-A for patients with soft signs of injury; and observation for completely asymptomatic patients.

Zone 3 injuries may involve the petrous or cavernous portions of the internal carotid artery, the vertebral artery or deep branches of the external carotid artery. Wide exposure of this area and distal vascular control may be difficult to obtain. Routine exploration as a diagnostic manoeuvre may be both hazardous and inaccurate; because of this, arteriography is the diagnostic method of choice for zone 3 injuries. Similar to zone 1 injuries, arteriography is sensitive and specific, and can be therapeutic for these relatively inaccessible injuries. Additionally, the intracranial circulation can also be assessed in patients with dense neurological deficits.

Specific injuries

Vascular

The repair of most arterial injuries can be accomplished by lateral arteriorrhaphy, primary reanastomosis or graft interposition. Ligation is rarely indicated for major vessels except in cases of inaccessible lesions, or in damage control situations. Vertebral artery ligation is generally tolerated without sequelae; ligation of the carotid is also generally tolerated in patients with an intact circle of Willis. Injury to major veins should be repaired if possible, although ligation does not lead to significant morbidity. The risk of venous air embolism, however, is of great concern. Suspected venous injuries should be managed with gentle pressure on the wound and exclusion of the injury with proximal and distal control as soon as possible. The patient should be in the Trendelenburg or at the least supine position until repair or ligation are complete to increase vascular pressure and reduce the risk of air embolus.

Tracheal

Most tracheal injuries can be closed primarily; mucosal approximation should be performed carefully with fine absorbable sutures to avoid intraluminal granulation tissue. Injuries of the thyroid or cricoid cartilage can be closed with sutures or plates. Tracheal rings and cartilage can be reapproximated using non-absorbable extramucosal sutures. All repairs should be buttressed with healthy tissue; this is of paramount importance if there is a concomitant oesophageal injury, as prevention of a tracheo-oesophageal fistula is necessary. Buttressing can be done with one of the strap muscles, or the sternocleidomastoid.

Oesophageal

Intraoperative oesophagoscopy and insufflation of a nasogastric tube with air or dye can aid in the diagnosis of suspected oesophageal injury. Oesophageal injuries should be widely explored to ensure the entire mucosal defect is seen. Once the full extent of injury has been identified it can be repaired in one or two layers with absorbable suture. As with tracheal injuries, all oesophageal repairs should be buttressed with healthy tissue, and wide drainage is highly recommended. If there has been significant delay in treatment or severe tissue destruction, the repair can be protected with a T-tube, or diversion can be performed with a proximal cervical oesophagostomy. More distal oesophageal injuries may be primarily repaired if diagnosed early, but late repair of intrathoracic oesophageal injuries should not be attempted. These injuries are best treated by wide drainage and oesophageal diversion and exclusion. Injuries in the thoracic and abdominal oesophagus can freely contaminate the peritoneal, mediastinal and thoracic cavities, leading to sepsis and death in 20–42% of cases. In contrast, mortality for cervical perforations ranges from 0% to 20%, with the majority of studies demonstrating zero mortality.

Conservative management is an acceptable strategy in selected patients. Indications for non-operative management include contained perforation, absence of distal obstruction, cervical or thoracic location, and minimal systemic signs of infection. Medical management consists of fasting, antibiotics and strict monitoring. Follow-up imaging is performed in 7–10 days to assess for healing

Blunt neck injuries

Blunt neck injuries are caused frequently by a direct blow to the thyroid or cricoid cartilage as the result of either motor vehicle or 'clothes-line' accidents. Severe injuries may also follow acceleration–deceleration forces, significant flexion–extension of the cervical spine or inadequate use of seat belts. Injuries to the larynx, trachea or oesophagus may result. Severe injuries may not be obvious on initial examination. Patients may present with airway obstruction, hoarseness, dysphagia, subcutaneous air, laryngeal deformities, blood in the aerodigestive tracts, haematoma or soft tissue swelling. Airway obstruction usually requires placement of a surgical airway. In the presence of severe laryngotracheal trauma a cricothyroidotomy often is not feasible, and tracheostomy will be required. In stable patients the first-line diagnostic test is a CT scan of the neck with contrast. The CT will allow adequate visualization of the skeletal, soft-tissue and vascular structures of the neck. The weakness of CT is its lack of sensitivity for aerodigestive injury; if suspicious physical examination or CT findings are present, bronchoscopy with oesophagoscopy or oesophagography may be indicated.

Signs of major vascular injury from blunt trauma may be more subtle than those associated with penetrating trauma. Injuries are frequently missed, or present in a delayed manner with a developing neurological deficit; however, blunt cerebrovascular injury is rare even in centres with aggressive screening policies. Injuries can occur as a result of severe hyperextension and rotation, direct force to the vessel or by injury from adjacent fractures. The majority of patients have minimal to no symptoms at presentation, but the risk of subsequent stroke may be high in the absence of treatment. Indications for screening in asymptomatic patients include seat belt sign, GCS score >8, presence of diffuse axonal injury, basilar skull fracture, cervical spine fracture, Le Fort II or III fractures, near hanging injury, direct blow to the neck and hyperextension injuries. Angiography has been the historic gold standard for diagnosis of blunt cerebrovascular injury; however, CT-A with multichannel (16–64) detectors is fast becoming the primary screening and diagnostic modality. CT-A has the benefit of being quick, widely available and easy to perform, and does not require the presence of a specially trained angiography team. It also avoids the morbidity associated with an arterial puncture and allows adequate visualization of the cervical spine and soft tissues. Studies have shown detection rates similar to angiography. Sensitivity ranges from 83% to 100%, with negative predictive values of 92–98%.

Injuries are graded using the Denver Grading Scale. Grade I injuries are small intimal tears or haematomas affecting <25% of the vessel lumen. Grade II injuries are intimal disruptions affecting ≥25% of the lumen. Grade III injuries are pseudoaneurysms. Grade IV injuries are complete vessel occlusions and grade V are vessel transections. The optimal type and duration of treatment for blunt cerebrovascular injury remains controversial; options include anticoagulation, antiplatelet therapy, surgical repair and endovascular coils or stents. Grade I/II injuries have a low risk of bleeding and treatment consists of anticoagulation or antiplatelet agents. Both methods appear to have equivalent efficacy. The majority of these low-grade injuries will heal, and repeat imaging is indicated as antithrombotic therapy can be discontinued after healing. Grade III injuries are less likely to resolve with antithrombotic therapy alone, and, although the risk of rupture is low, the risk of thromboembolic complications without treatment can be high. Surgery should be considered for all accessible lesions. Endovascular therapies are available and placement of stents or coils may decrease the risk of enlargement; however, there is no evidence that they are superior to antithrombotic agents alone in stroke prevention. Stents used for carotid artery injuries in particular have been associated with greater complications than antiplatelet/anticoagulant therapy, and are not currently recommended. Grade IV lesions (occlusions) are unlikely to resolve with antithrombotic therapy, and neither medical nor surgical treatment is likely to reverse symptoms if already present. Follow-up imaging is not recommended for grade IV lesions as they are unlikely to change with time. Grade V lesions (active extravasation) require surgery or endovascular intervention.

Complications of cervical trauma

Massive haemorrhage leading to hypovolaemic cardiac arrest and complications arising from spinal cord injuries are the leading causes of mortality in patients with cervical trauma. Missed injuries, involving the oesophagus, may lead to fistula formation, abscess and occasionally to mediastinitis. Undiagnosed vascular injuries have produced late sequelae, including arteriovenous fistula and false aneurysms. Glottic or subglottic stenosis may complicate laryngeal and tracheal injuries.

Chest trauma

Anatomy

The chest and its contents can be divided into four anatomical zones: the chest wall, the pleural space, the lung parenchyma and the mediastinum. The chest wall consists of the bony thorax and the associated musculature. Injuries to the chest wall consist primarily of fractures of the ribs and sternum as well as soft-tissue injuries. The pleural space is a virtual space between the lung and the chest wall. Although it is normally empty, the pleural space can become filled with air (pneumothorax) or blood (haemothorax) as a result of injury. In addition, alteration of the normally negative intrapleural pressure will cause collapse of the lung and ventilatory compromise. The lung parenchyma and airways can become compromised by laceration, contusion, obstruction, haematoma and pneumatocele. The mediastinum contains the heart, aorta, great vessels, trachea, proximal bronchi and oesophagus. Injuries within the mediastinum have the potential to involve many critical structures, and require an aggressive diagnostic and therapeutic approach.

Initial management

The initial management of patients with suspected chest injury should follow the ABCDEs; a primary survey should be done to identify immediate life-threatening injuries. As mentioned above the major thoracic injuries that should be identified and addressed during the primary survey include tension pneumothorax, massive haemothorax, open pneumothorax and flail chest in the breathing section; and pericardial tamponade in the circulatory section.

Tension pneumothorax and massive haemothorax may both present with haemodynamic compromise and absence or diminished breath sounds on the affected side, and tracheal deviation away from the side of injury. Tension pneumothorax/haemothorax and open pneumothorax should all be treated with immediate chest tube placement. Pericardial tamponade may also present with profound hypotension, distended neck veins and muffled heart sounds, and is generally the result of penetrating trauma to the cardiac box. This box is bordered by the clavicles and costal margins superiorly and inferiorly, and bilaterally at the mid-clavicular line. Diagnosis can be confirmed with pericardial views on FAST, although if the bloody effusion decompresses into the left chest a false-negative FAST is possible. For this reason negative FAST cannot definitively rule out pericardial effusion in the presence of a left haemothorax. Rapid thoracotomy or median sternotomy to release the tamponade and repair the cardiac defect are life saving and should be performed immediately.

Once the primary survey is complete, the secondary survey should focus on visual inspection of the chest for deformities, contusions, lacerations and penetrating wounds. Radio-opaque markers should be placed at all wound sites prior to imaging. The location and direction of penetrating injuries can aid in the determination of structures at risk for injury. However, it is important to be aware that structures within the chest may be damaged by penetrating wounds outside the boundaries of the thorax.

Palpation of the chest can reveal areas of tenderness and deformity as well as crepitance. These may be indicative of rib or sternal fractures. Marked crepitance is indicative of subcutaneous emphysema, which can result from major bronchial injury, pulmonary parenchymal injury or oesophageal injury. Auscultation of the chest allows assessment of breath sounds; decreased breath sounds are suggestive of pneumothorax/haemothorax, or possible mainstem bronchus intubation.

The CXR can yield a wealth of useful diagnostic information and should be obtained as early as possible in the resuscitation. The radiograph should be examined carefully in its entirety for evaluation of the bony thorax, soft tissues of the chest wall and mediastinum, lung parenchyma, airways, and cardiac and mediastinal silhouette. It should also be repeated if interventions occur such as intubation or nasogastric or thoracostomy tube or central line placement, or if the clinical course alters significantly. An arterial blood gas should be performed as well to assess for hypoxia, hypercarbia, acid–base status and toxins such as methaemoglobinaemia or carboxyhaemoglobin. Other more involved diagnostic studies of the chest are required in special circumstances. These may include CT-A, angiography, FAST ultrasound and formal echocardiogram. These may be necessary for the diagnosis of thoracic or great vessel injury, thoracic spinal fractures and blunt cardiac injury.

Injuries to the chest wall

Rib fracture

By far the most common injury sustained after blunt chest trauma is rib fracture (Figure 14.12). In older patients, rib fractures can occur after a relatively minor injury because of decreased bone density and loss of chest wall compliance. The primary manifestation of rib fractures is pain. Pain can cause significant complications owing to splinting and poor inspiratory effort, including alveolar derecruitment, hypoxia, respiratory distress, airway obstruction and pneumonia. The use of bedside pulmonary function tests including forced vital capacity (FVC) and maximum inspiratory force can give an objective estimate of the patient's degree of ventilator compromise. Tidal volume of $<5\,\mathrm{mL/kg}$, FVC $<10\,\mathrm{mL/kg}$ or maximum inspiratory force of $<30\,\mathrm{cmH_2O}$ are indicators of ventilatory compromise. Additionally non-invasive continuous capnography, which can be used to follow the trend of expired carbon dioxide, can help to detect hypoventilation early.

Initial attempts at pain control using oral or parenteral analgesics are appropriate, but additional measures must be aggressively pursued in those patients who do not respond to initial therapy. Adjuncts can include topical pain patches with local anaesthetic, intercostal nerve blocks, and epidural or spinal anaesthesia.

Rib fractures may also serve as a marker for more serious injuries. Historically, fracture of the first rib has been a hallmark of very-high-energy transfer and has been associated with major chest, abdominal and vascular injury. Fractures of the lower ribs are associated with solid organ injury, with a 20% incidence of splenic injury associated with fracture of the tenth to twelfth ribs on the left side.

Flail chest

Flail chest occurs when three or more contiguous ribs are fractured in two or more locations resulting in a segment of chest wall that can move independently (Figure 14.13). This will result in paradoxical motion with negative-pressure ventilation. Additionally, a large amount of force is required to create a flail segment and may also cause underlying pulmonary contusion. Many physiological changes occur with a flail chest: the paradoxical motion can decrease total lung capacity and functional residual capacity, pain associated with rib fractures can result in splinting and atelectasis, and most importantly the associated contusion can lead to ventilation–perfusion mismatch. These changes ultimately culminate in increased shunt fraction, resulting in hypoxia. While the mechanical injury to the chest wall undoubtedly contributes to respiratory morbidity, the underlying pulmonary contusion is by far the most important determinant of respiratory status. The majority of patients do not require mechanical ventilation, but, when necessary, mechanical ventilation is associated with worse prognosis and a variety of complications. The need for intubation and mechanical ventilation is determined by the patient's clinical appearance, ability to ventilate adequately and the degree of pulmonary dysfunction. Pain control and pulmonary toilet are essential in preventing complications associated with flail chest. Several studies have demonstrated the superiority of epidural anaesthesia for pain control, prevention of pneumonia and decreased need for mechanical ventilation. Increasing age is associated with worsened outcomes, with some studies suggesting that ages as low as 45 years are associated with increased complications and mortality. Thus adequate analgesia and pulmonary toilet are of paramount importance in the management of the elderly patient with thoracic trauma and strong consideration should be given to ICU admission.

Some patients with large flail segments have very severe chest wall deformity and subsequent loss of volume in the affected hemithorax. Chest wall reconstruction and mechanical stabilization of the flail segment is advocated by some for highly selected cases. There is some evidence to suggest that plating, wires or other means of mechanical stabilization may contribute to weaning from mechanical ventilation and pain control. However, evidence is limited and definite indications for surgical stabilization are lacking.

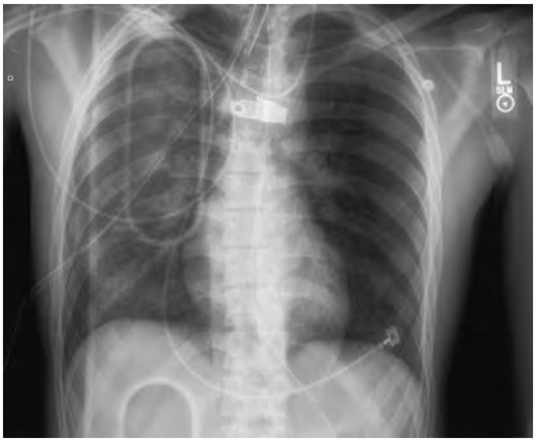


Figure 14.12 Chest radiograph demonstrating multiple posterior rib fractures of the right chest.

Sternal fracture

The majority of sternal fractures (Figure 14.14) are sustained as a result of the use of shoulder restraints, especially in older patients with decreased bone density. Under these circumstances, the sternal fracture is often an isolated injury and is rarely associated

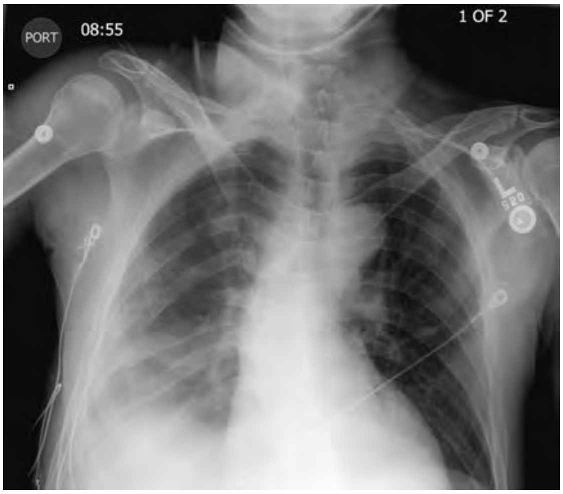


Figure 14.13 Chest radiograph demonstrating a posterior flail segment of the right chest with associated haemothorax and right lower lobe contusion.

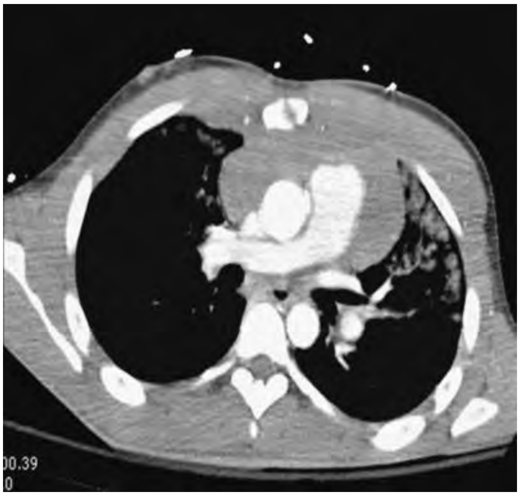


Figure 14.14 Axial CT images of a sternal fracture with associated mediastinal haematoma.

with significant intrathoracic injury. The need for additional evaluation in patients with sternal fracture should be based on the overall circumstances, not the mere existence of the fracture. Treatment of the sternal fracture itself is generally conservative with an emphasis on pain relief. Fewer than 25% of patients with a complete fracture of the sternum will require operative fixation.

Injuries to the pleural space

Pneumothorax

Pneumothoraces occur when air escapes from the injured lung, or enters through a penetrating wound and collects in the pleural cavity (Figure 14.15). Pneumothorax can be either simple or present with tension physiology. Tension pneumothorax is the most rapidly life threatening of breathing problems. It occurs when air continues to accumulate in the pleural space, creating pressures above atmospheric. This can result in mediastinal shift away from the site of injury and cause markedly decreased venous return to the heart and resultant hypotension. Tension pneumothorax can be recognized by hypoxia, tachypnoea, hyper-resonance, absence of breath sounds on the side of injury and deviation of the trachea away from the side of the injury, distended neck veins and hypotension. If tension pneumothorax is suspected emergency decompression must be performed. ATLS recommends needle decompression with a large-bore needle placed in the second intercostal space in the mid-clavicular line. However, placement of a needle in the fifth or sixth intercostal space at the anterior axillary line corresponding to the level of the inframammary crease is frequently easier and further removed from vital mediastinal structures. Needle decompression should be followed by the placement of a chest tube for more permanent decompression of residual air and drainage of any blood that may be associated with the now simple pneumothorax. Simple pneumothoraces are generally successfully treated with placement of tube thoracostomy as well. Occult pneumothoraces (seen on CT scan but not on CXR) or small asymptomatic pneumothoraces may be observed and can resolve spontaneously. However, expectant management should not be pursued in patients on positive-pressure ventilation as the risks of expansion and creation of tension physiology are significantly elevated, and the time required for air leaks to seal can be greatly prolonged.

Open pneumothorax, or sucking chest wounds, occur when a defect in the chest wall is full thickness and large enough for air to communicate between the thoracic cavity and the environment. If the wound is sufficiently large, intrapleural pressure remains equal to atmospheric pressure. Attempts at spontaneous ventilation will result only in movement of air in and out of the body wall defect. No significant ventilation of the lung is possible and respiratory compromise is severe. The wound should be grossly decontaminated and an occlusive dressing should be applied and secured on three sides. A chest tube should be placed to relieve the associated pneumothorax immediately.

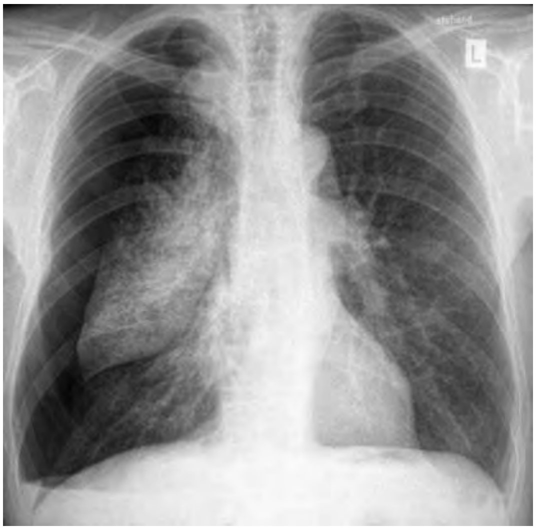


Figure 14.15 Chest radiograph demonstrating right simple pneumothorax; note that the mediastinal structures remain in the midline despite significant lobar collapse.

Haemothorax

Massive haemothorax may also present with tension physiology. Hypotension may be a result of decreased preload from tension physiology as well as from massive blood loss. Distended neck veins may or may not be present, and obstruction due to tension physiology increases venous pressure causing distension; however, severe associated intravascular volume depletion may counteract this effect. Simple haemothorax can present with diminished breath sounds, dullness to percussion and effusion or blurring of costophrenic angles on CXR (Figure 14.16). Treatment for both massive and simple haemothorax is immediate placement of a chest tube. The majority of lacerations to the lung involve low-pressure vessels of the pulmonary circulation and bleeding can be expected to stop after placement of the chest tube and re-expansion of the lung. Injuries to the intercostals or internal

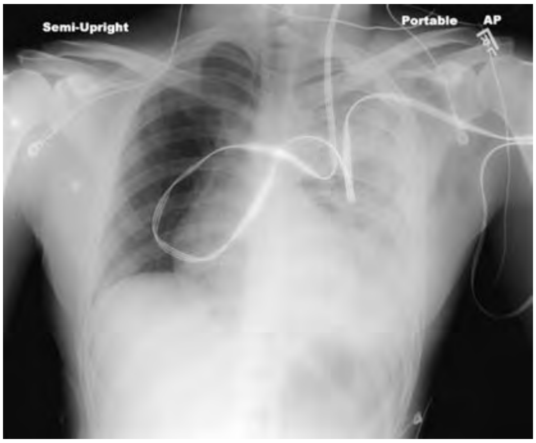


Figure 14.16 Chest radiograph demonstrating left haemothorax with near white-out of the left chest; this is typical of images taken with the patient supine as the blood layers evenly over the posterior thoracic wall.

mammary arteries, however, can lead to massive haemorrhage, especially if the vessels are only partially transected. Blood loss >1000–1500 mL defines a massive haemothorax and is an indication for operative exploration. Additional indications for thoracotomy include continued blood loss of ≥200 mL/h for longer than 4 hours and massive air leak. Every effort should be made to collect shed blood in a sterile fashion so that it may be autotransfused.

Retained haemothorax can complicate the hospital course of 3–8% of patients with haemothorax. It can occur as a result of delay in presentation, diagnosis or treatment; thoracostomy tube malposition, migration or occlusion; and may result in empyema or fibrothorax. Treatment options include open drainage, video-assisted thoracoscopic drainage and intrapleural thrombolysis. Intrapleural thrombolysis can lead to radiographic resolution in up to 90% of cases.

Technique of tube thoracostomy

The initial therapy in most patients with haemothorax/pneumothorax is tube thoracostomy. The tube should ideally be placed rapidly but in a controlled and sterile fashion. The arm is abducted if possible, the chest is rapidly prepared with Betadine or chlorhexidine and the insertion site is anaesthetized if time permits and the patient is awake. The ideal insertion site is in the fourth or fifth intercostal space along the mid-axillary line. This generally corresponds to the inframammary fold in females and the nipple in males. This site is chosen to avoid injury to the diaphragm and to allow easiest access to the thoracic cavity through a minimal amount of muscle and soft tissue. It is important to palpate the entry site carefully to ensure no injury to the diaphragm has occurred, that the thoracic cavity has been entered and that there are no significant adhesions in the area. Entry should meet no resistance, and the tube is generally advanced 10–15 cm such that the last drainage hole is within the thorax. The tube should then be securely sutured and taped in place with an occlusive dressing and attached to a closed drainage system that has been prepared for autotransfusion. A postprocedure CXR is mandatory both to confirm placement of the tube and to ensure the desired therapeutic effect. Persistent haemothorax/pneumothorax (Figure 14.17) should be treated with a second chest tube, but should also raise the question of the need for a thoracotomy to control bleeding or address airway injury.

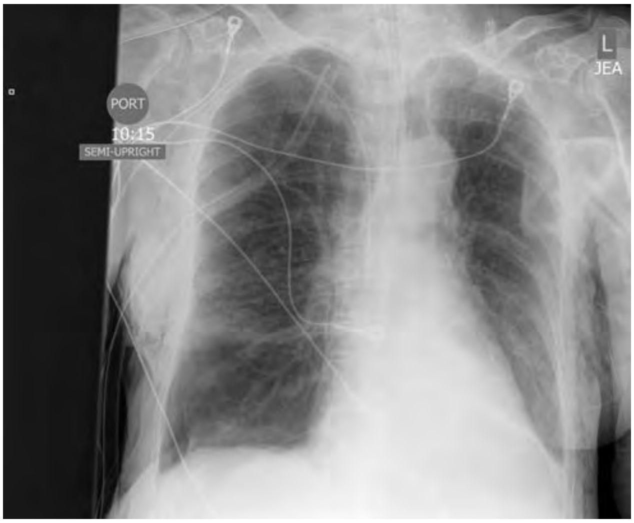


Figure 14.17 Chest radiograph demonstrating small residual subpulmonic pneumothorax after chest tube placement.

Pulmonary parenchymal injuries

Pulmonary laceration

Simple lacerations of the lung parenchyma are common after penetrating trauma but uncommon after blunt trauma. Laceration can result in leakage of air or bleeding, resulting in a pneumothorax/haemothorax (Figure 14.18). Initial treatment of either is with a tube thoracostomy to drain any air or blood and re-expand the lung. The majority of bleeding will cease spontaneously as the lung vasculature is a low-pressure system, and the lung is rich in tissue thromboplastin. The majority of air leaks are small and will close spontaneously. Chest tube placement is the definitive management in the majority of cases.

Air leak may be prolonged or exacerbated by positive-pressure ventilation. The treatment is generally weaning from mechanical ventilation and extubation, which generally allows healing and resolution of the leak. In the case of massive air leaks tracheobronchial injury should be investigated. If the air leak is large enough effective alveolar ventilation can be lost. If this is the case selective bronchial intubation beyond the injury, or of the unaffected lung, can be used to decrease leakage and improve ventilation. In extreme cases split lung ventilation with a double-lumen ETT and two separate ventilators can be used to allow independent lung ventilation.

Parenchymal injury requiring extensive lung resection and pneumonectomy following trauma is rare but can result in severe cardiac dysfunction and high mortality. In fact morbidity and mortality increase in a stepwise fashion as the extent of pulmonary resection increases. Reasons for this include increasing injury severity and more associated injuries. However, several studies have shown that the extent of resection remains an independent predictor of mortality even after correcting for

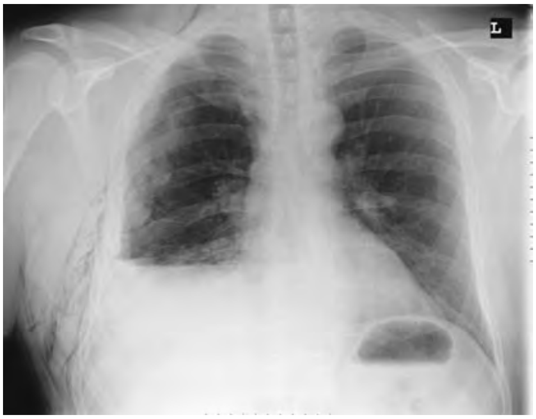


Figure 14.18 Chest radiograph demonstrating a right haemopneumothorax due to a stab wound. Note the air–fluid level in the lower chest and the associated subcutaneous emphysema.

these factors. This may be due to increased right heart strain/failure following traumatic lung resection. In contrast to patients undergoing lung resection for non-traumatic reasons, e.g. when chronic lung disease has resulted in minimal blood flow to the diseased lung preoperatively, trauma patients generally have normal pulmonary vasculature. After resection the sudden decrease in the pulmonary vascular compliance owing to parenchymal resection acutely increases right heart afterload causing strain and failure. As the right heart fails there is decreased left heart preload, resulting in decreased cardiac output. There are case reports using pulmonary vasodilating agents such as nitroglycerin, nitroprusside, alprostadil, epoprostenol and nitric oxide to treat right heart failure following traumatic resection and pneumonectomy. However, treatment is primarily supportive, and mortality remains high: 62% following pneumonectomy, 35% following lobectomy and 22% following wedge resection.

Pulmonary contusion

Pulmonary contusion is a common finding following chest trauma. Contusions occur most commonly after blunt thoracic trauma, with falls and motor vehicle crashes accounting for the majority of injuries in the civilian population. Among military combatants the shock wave produced by high-velocity weapons and blasts also commonly results in severe pulmonary contusion. Contusion results from haemorrhage and oedema formation in the injured lung parenchyma. Contusion can result in decreased lung compliance, diminished alveolar function, vasoconstriction in the injured lung, increased secretion production, and decreased secretion clearance. A significant shunt can occur and result in significant hypoxia. Decreased compliance can also increase the patient's work of breathing, causing respiratory fatigue and failure and increased ventilator support requirements in the intubated patient. The diagnosis of pulmonary contusion is generally made on the basis of mechanism of injury, arterial hypoxaemia and CXR or CT scan findings. Parenchymal damage generally becomes radiographically apparent within 4–6 hours of injury, peaks at approximately 72 hours after injury and usually resolves within 7 days. However, the period of time needed for return of adequate lung function may vary from 1 day to several weeks depending on the severity of injury, inflammatory response, concomitant injury and comorbid conditions, as well as the presence or absence of complications such as ARDS and pneumonia.

Plain films and CT scans are the mainstay of diagnosis (Figure 14.19a,b). CXR may demonstrate an ill-defined area of infiltration in the area of injury; however, it may take some time for the complete extent of injury to manifest, with only 47% of lesions detected at admission and 92% present at 24 hours. CT scan may be more sensitive than plain radiographs, particularly in the initial period following injury, in delineating the extent of contused lung. Scoring systems aimed at predicting the extent of pulmonary dysfunction and contusion-related complications are generally more accurate when using CT findings than when using plain radiographs. However, newer generation CT scanners may be overly sensitive and differentiating true contusion from atelectasis, aspirated contents and haemothorax may be difficult; therefore, diagnosis should be made by correlating radiographic findings with clinical symptoms. Lesions deemed 'occult contusions', i.e. those seen on CT but not visible on CXR, are generally asymptomatic and self-limited.

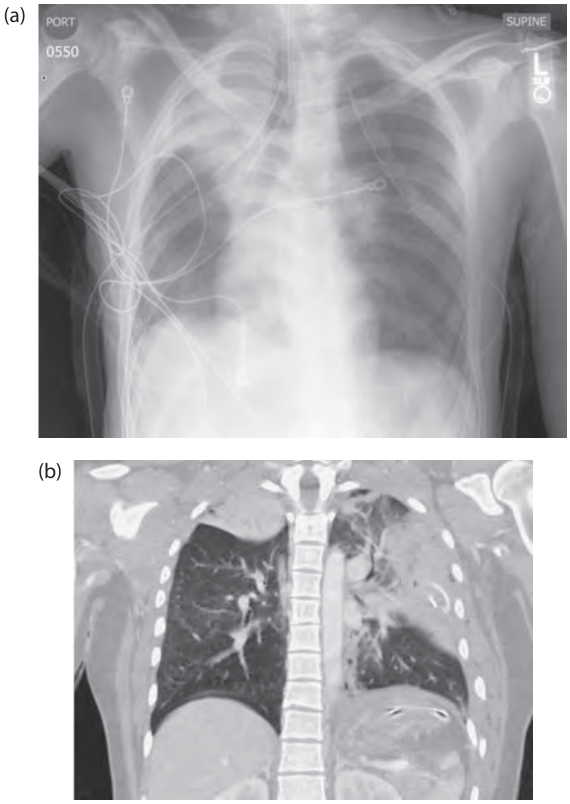


Figure 14.19 (a) Chest radiograph demonstrating a right upper lobe contusion. (b) Coronal CT demonstrating bilateral lung contusions.

The treatment of pulmonary contusions is mainly supportive with supplemental oxygen and aggressive pulmonary toilet. In selected patients non-invasive positive-pressure ventilation may be helpful in improving oxygenation, and avoiding intubation in up to 82% of patients; however, any patients who manifest signs of respiratory failure should be intubated immediately. For patients with severe hypoxaemia due to unilateral lung injury the use of a rotating bed or dependent positioning of the uninjured lung may improve ventilation–perfusion mismatch and improve oxygenation. The use of rotating beds has also been associated with decreases in ventilator days and associated pneumonias. Decreased capillary endothelial integrity in the injured lung makes judicious fluid management in the patient with pulmonary contusion essential. There is some evidence to suggest a role for replacement of surfactant in patients with severe contusions; however, there is no role for prophylactic antibiotics or steroids in the treatment of pulmonary contusions.

Injury to the large airways

The incidence of major tracheobronchial injury in most trauma centres is low. This is the result of a combination of high scene mortality and a tendency for some injuries to remain relatively asymptomatic in the acute period. In general, penetrating injuries occur in the cervical area, while the majority of blunt injuries occur within 2.5 cm of the carina. Patients present in one of two ways, depending on the location of the bronchial rupture.

Patients in whom the bronchus ruptures into the pleural cavity will present with pneumothorax and air leak after chest tube placement. Other symptoms can include massive subcutaneous or mediastinal emphysema and haemoptysis. If the bronchial rupture occurs within the mediastinum the presentation can be much less acute. Respiratory distress in this group is uncommon; symptoms may include haemoptysis, mediastinal emphysema (Figure 14.20) and occasionally subcutaneous emphysema. These types of injury generally heal without intervention; however, they may present in a delayed fashion with symptoms due to strictures in the area of injury.

Patients with suspected injury should be evaluated with bronchoscopy. Treatment depends on the extent and location of the injury as well as the patient's overall clinical condition. Patients with concomitant oesophageal injury should all undergo repair. Patients with persistent symptoms following adequate pleural drainage and intubation distal to the site of injury may require repair. Injuries involving the membranous portion of the trachea are more likely to heal than injuries involving the cartilaginous portion. Injuries involving less than one-third of the circumference of the trachea or major bronchus can be observed. Injuries greater than one-third require repair to prevent stricture formation.

Injuries to mediastinal structures

Blunt thoracic aortic injuries

Blunt thoracic aortic injury is primarily the result of sudden deceleration injuries, and injuries primarily occur in areas where the aorta is fixed. The most common site of injury is in the descending aorta just distal to the origin of the left subclavian at the site of the ligamentum arteriosum; this site accounts for >90% of injuries. Other locations include the aortic root, where the aorta is fixed to the heart, and the diaphragmatic hiatus. Unfortunately, the majority of patients with thoracic aortic injury do not reach the hospital alive: 80–90% die prior to hospital arrival. Of those patients surviving to evaluation, thoracic aortic injury is present in approximately 0.3–2.8%. Patients generally fall into one of two categories. The first group presents with haemodynamic instability due to free rupture of the haematoma and time to treatment is a key factor in outcome: 30% die within 6 hours, 50% within 24 hours and 90% at 10 weeks if thoracic aortic injury is not treated. The second group consists of patients who are haemodynamically stable, and often asymptomatic; these patients generally have small contained pseudoaneurysms (Figure 14.21) with a low risk of rupture and will respond well to medical management and delayed treatment. However, because these patients are generally asymptomatic, screening and diagnosis of thoracic aortic injury may be difficult. Currently, CXR is the primary screening examination. Abnormalities associated with thoracic aortic injury include widened mediastinum (>8 cm), apical capping, left pleural effusion, loss of the aortopulmonary window, depression of the left mainstem bronchus, deviation of the trachea or nasogastric tube to the right, and widening of the right paratracheal stripe. However, CXR is non-specific, and all patients with abnormalities on CXR should rapidly undergo further imaging to definitively diagnose their injuries. Options include traditional angiography, CT-A, echocardiography and MRI/magnetic resonance angiography (MRA).

While specificity is excellent, ranging from 98% to 100%, the sensitivity of aortography ranges widely, from 38% to 92%. The majority of missed injuries appear to be intimal tears and small intramural haematomas, which are more likely to be diagnosed by echocardiography or CT-A. Access site complications include thrombosis, haemorrhage, dissection, pseudoaneurysm and arteriovenous fistula. Systemic complications include anaphylaxis and contrast-induced nephropathy, which can be as high as 62% in high-risk populations.

In contrast to angiography, CT-A is widely available, can be performed at any time, is fast and avoids arterial puncture. Similar to the use of CT-A in blunt cerebrovascular injury, CT-A is increasingly replacing traditional angiography for the diagnosis of thoracic aortic injury. A comparison of trauma practices

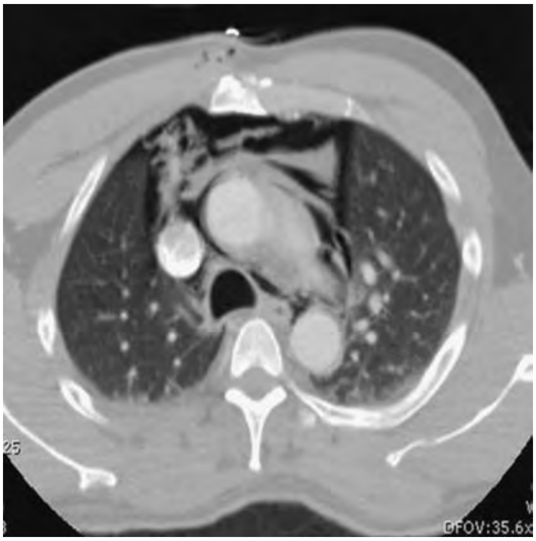


Figure 14.20 Axial CT scan demonstrating pneumomediastinum.

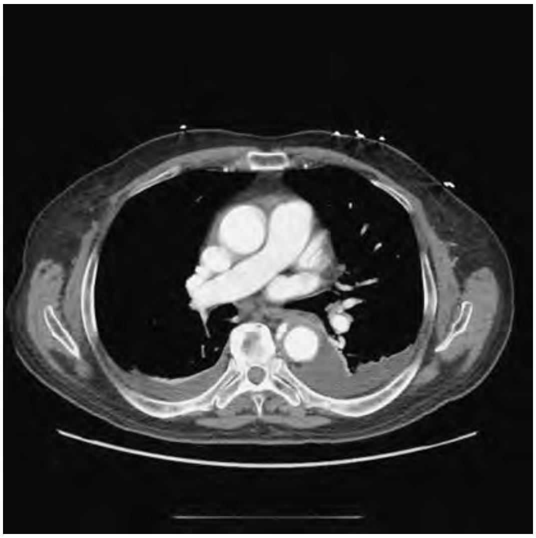


Figure 14.21 Axial CT scan with thoracic aortic pseudoaneurysm; note the associated contained haematoma.

from the time periods 1994–6 and 2005–7 revealed an increase in the use of CT scan for diagnosis of thoracic aortic injury from 34.8% to 93.3%. Modern series reliably report sensitivity between 95% and 100%. Specificity with modern technology ranges between 94% and 99.8%.

In patients with haemodynamic instability or who require emergency surgery, transoesophageal echocardiography may be the ideal diagnostic modality. It is portable, fast and repeatable, and does not require intravenous contrast. Sensitivity can be as high as 98–100%, but can vary widely based on operator experience. An alternative to transoesophageal echocardiography is intravascular ultrasound. This technique involves insertion of an ultrasound probe intravascularly into the aorta. It has excellent diagnostic accuracy, with sensitivity of 92–100% and specificity nearing 100%. In a study comparing intravascular ultrasound, transoesophageal echocardiography and angiography among patients with blunt chest trauma, intravascular ultrasound outperformed both transoesophageal echocardiography and angiography. Drawbacks of transoesophageal echocardiography include invasiveness and the need for specialized equipment. There are also risks of oesophageal injury and airway compromise. Intravascular ultrasound requires arterial puncture with the attendant risks of access site complications such as haematoma, fistula, pseudoaneurysm and dissection.

Contrast-enhanced MRA (CE-MRA) is performed with intravenous gadolinium and produces images of a high quality with fewer respiratory and cardiac artefacts than traditional MRI. CE-MRA can be used for diagnosis with a sensitivity and specificity between 98% and 100%. CE-MRA has the ability to create detailed three-dimensional reconstructions that are of great value in preoperative or prestent planning. CE-MRA has the benefit of avoiding ionizing radiation and iodinated contrast. However, CE-MRA requires lengthy examination times outside of a monitored critical care setting, and potentially lengthy transport times. CE-MRA may not be available in many centres, and may not be operational at night, on holidays or weekends. Because of this, it is still ill-suited to the acute trauma work-up.

Patients with thoracic aortic injury historically all underwent emergency open repair via a left thoracotomy. Recently, conservative management with blood pressure and heart rate control have allowed many patients to be temporized until open or endovascular treatment can be undertaken after other life-threatening injuries are addressed and normal physiology has been restored. Medical management consists of tight blood pressure and heart rate control with beta-blockers and vasodilators. In cases of minor injury and in patients with prohibitive operative risk, medical management may be the definitive treatment. However, the majority of patients undergo either open or endovascular repair. Open repair is associated with significant mortality and morbidity, the most significant of which is paraplegia, occurring in 1.6–14.0% of patients. Because of the difficulties and dangers of open repair the use of endovascular techniques has significantly increased in popularity. There have been several studies of thoracic endovascular aortic repair demonstrating a technical success rate of 92–100% with mortality of 0–17.9%. In comparison with open repair, studies have shown a significant decrease in mortality, blood loss, transfusion requirements, paraplegia rates, operating theatre times and length of stay. Complications, including access site dissection, fistula and pseudoaneurysm, and systemic, such as CIN, pneumonia, heart attack and stroke, occur in 0–15% of cases. Endoleaks also occur in 0–15% of patients and may require additional intervention. The majority are type I endoleaks; this is likely to be related to the technical difficulty in achieving a good proximal seal in the aortic arch. The majority of type I leaks are diagnosed at the time of the original procedure and treated. However, long-term outcomes are unknown and several studies have demonstrated an increased need for secondary interventions among thoracic endovascular aortic repair patients compared with those undergoing open repair.

Penetrating thoracic aortic injuries

Injuries to the thoracic aorta and great vessels are more commonly due to penetrating mechanisms. In these circumstances the diagnostic problems are considerably more clear-cut. Those patients presenting with initial haemodynamic instability and a penetrating wound to the chest require urgent thoracotomy. In stable patients CT has proven to be an excellent and cost-effective screening examination. Patients with transthoracic or transmediastinal gunshot wounds are at risk for vascular, aerodigestive and spinal injuries. CT scan can determine which patients require surgery and which require additional more invasive imaging, such as angiography, oesophagoscopy, oesophagography or bronchoscopy. Studies reveal conclusive results in 82.9–100% of cases.

The approach to major vascular injuries in the thorax may involve significant blood loss; therefore, adequate preparations should be made, including initiation of the MTP and having large-bore intravascular access and warming equipment available. Proximal and distal control may require cross-clamping of the descending aorta, with the associated risk of spinal cord ischaemia. Partial bypass techniques have greatly improved over the past several years and should be used within the trauma setting whenever practical. Under most circumstances, systemic heparinization is not required. Primary repair of the injury should be attempted when feasible, but intercostal vessels should not be sacrificed to gain mobility owing to the increased risk of spinal cord ischaemia. Dacron interposition grafts can be used if defects are too large for primary repair. In patients with severe physiological derangement, temporary intravascular shunts may be used as a damage control technique. Chest tubes, nasogastric tubes and commercially prepared vascular shunts can all be used depending on the diameter of the injured vessel. The shunts should be secured proximally and distally with silk ties or umbilical tape. Patency appears to be adequate without the use of systemic heparinization, although maximum dwell times are unknown.

Blunt cardiac injuries

The incidence of blunt cardiac injury varies widely from 8% to 71% owing primarily to the wide spectrum of the disease, which ranges from mild asymptomatic contusion to free wall rupture, and a lack of consensus on a gold standard diagnostic test. Blunt cardiac injury can be divided into abnormalities of conduction and those of structure.

The most commonly seen electrical abnormalities are non-specific ST–T wave changes. The most common clinically significant electrical disorders are tachyarrhythmias, of which atrial fibrillation is the most likely. Haemodynamically significant arrhythmias are rare and treatment is similar to arrhythmias of non-traumatic aetiology with medications or cardioversion. There are no data to support the use of prophylactic antidysrhythmics.

Structural lesions include blunt cardiac rupture, valvular or papillary muscle rupture, coronary artery injury and myocardial contusion. Structural lesions are best diagnosed with echocardiography, which has an excellent sensitivity and specificity. Few patients with blunt cardiac rupture remain alive long enough to present to the hospital. The right chambers are most commonly injured because of their more vulnerable location directly beneath the sternum. If patients reach the hospital alive the most common presentation is haemorrhagic shock or tamponade. Treatment is swift diagnosis and open cardiac repair but subsequent mortality remains high. Patients with suspected cardiac rupture should be supported with volume resuscitation with early initiation of MTPs. Needle pericardiocentesis may be useful only as a temporizing measure if there are no immediate surgical capabilities. Valvular or papillary muscle rupture is rare after blunt injury but can be progressive, making prompt surgical intervention the most appropriate treatment. Blunt coronary artery injury is also very rare, representing <2% of blunt cardiac injuries. Direct impact may result in thrombosis, dissection or vessel rupture. Coronary rupture may present with shock, tamponade or haemorrhage; treatment is immediate operative repair, or ligation if tolerated. Occlusion or dissection present similarly to acute myocardial infarction. Treatment includes medical management, percutaneous intervention and operative revascularization.

Contusions are the most common structural abnormality. The majority of cardiac contusions are asymptomatic and have no long-term consequences. However, patients with haemodynamic instability should be admitted to the ICU, with intubation, invasive cardiac monitoring and urgent echocardiogram. Treatment of cardiogenic shock associated with blunt cardiac injury should include replacement of intravascular volume, inotropic agents and occasionally placement of an intra-aortic balloon pump. In stable patients with suspicion of blunt cardiac injury diagnostic tests include creatine phosphokinase-MB (CK-MB), troponin I (Tn-I), echocardiography and ECG. There is level I evidence to support the use of ECG in diagnosis of blunt cardiac injury, and it is recommended that all patients with suspicion of blunt cardiac injury undergo at least an admission ECG. If this is normal there is little risk of blunt cardiac injury. However, addition of biomarkers of cardiac injury such as CK-MB or Tn-I can increase sensitivity and specificity to 100% and 71%, respectively.

If blunt cardiac injury is diagnosed, patients should be admitted for observation. The majority of haemodynamically significant complications occur in the first 24–48 hours. If they are appropriately monitored and managed the vast majority of patients recover quickly with no long-term cardiac dysfunction. In the immediate period studies have shown that, if early surgery is required for other injuries, general anaesthesia can be used safely, although consideration should be given to intraoperative haemodynamic monitoring with transoesophageal echocardiography or pulmonary artery catheters.

Penetrating cardiac injuries

Penetrating cardiac trauma is rare, accounting for only 0.16% of trauma admissions overall. Even with modern trauma systems only 10–15% of victims of penetrating cardiac trauma reach the hospital alive. Stab wounds have a better outcome than gunshot wounds and account for the majority of survivors of penetrating cardiac injury. Predictors of good outcome include the presence of a sinus rhythm upon entry to the pericardium, absence of exsanguination and presence of tamponade. Negative predictors include emergency department thoracotomy and atrial injury (Figure 14.22), and wounds producing large defects in the pericardium. Lastly, longer prehospital times and patients requiring >3 minutes of CPR have higher mortality.

The highest risk for injury to the heart occurs within the cardiac box – the area bordered by the clavicles, costal margin and mid-clavicular lines. However, every penetrating injury to the chest should be suspected of cardiac injury until proven otherwise. Hypotension is the most common clinical finding; however, Beck's triad (hypotension, distended neck veins, distant cardiac sounds) may not be present because of significant intravascular volume depletion associated with haemorrhage. Anxiety, restlessness and pulsus paradoxicus are all also associated with tamponade.

In the unstable patient with multicavitary trauma, or in those with thoracic trauma who are stable, FAST has become the diagnostic method of choice. With reported sensitivity and specificity nearing 100% in diagnosing pericardial effusion, a positive cardiac FAST can be seen as an absolute indication for operation. However, false-negative results can occur, particularly in the presence of significant pneumothorax/haemothorax allowing decompression of a potential effusion into the left hemithorax. Given the widespread dissemination of FAST, invasive testing such as pericardiocentesis and subxiphoid

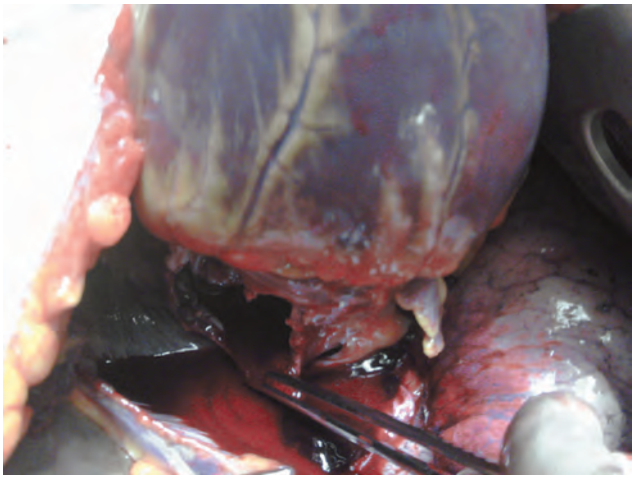


Figure 14.22 Postmortem photograph depicting a large right atrial laceration due to a gunshot wound to the chest.

window have few indications in a hospital environment and should only be considered when FAST is not available. In contrast, a transdiaphragmatic window during laparotomy for associated injuries is useful for rapidly ruling out cardiac injury. It is easily performed, has excellent sensitivity and specificity, and is associated with a very low rate of complications. If the pericardial fluid is bloody, the laparotomy can be extended into a median sternotomy to address the cardiac injury rapidly.

Keys to management include early recognition, aggressive resuscitation and rapid repair. Patients with a perfusing rhythm and positive FAST should be transferred to the operating room for median sternotomy immediately. However, many patients arrive *in extremis* or full cardiac arrest and there is no time for transfer to the operating theatre. These patients should undergo immediate resuscitative thoracotomy. Resuscitative thoracotomy has an overall survival rate of 7.4%; the best outcomes, with 19.4% survival, are seen with thoracic stab wounds. In the presence of cardiac arrest, the thoracic aorta should be clamped in order to improve coronary and cerebral circulation. Once the cardiac injury is identified, active bleeding is controlled by finger compression and then sutured. Injuries close to the coronary vessels should be repaired with horizontal mattress sutures underneath the vessel. Injuries to the posterior wall of the heart can be exposed by slowly lifting the apex of the heart. Overly rapid or aggressive elevation can result in severe arrhythmias or complete arrest. Following repair, if necessary, the heart is resuscitated with open cardiac massage, intravenous fluids, intravenous or intracardiac epinephrine, and internal defibrillation.

All survivors should undergo serial postoperative echocardiographic evaluation. Penetrating injuries may lacerate the septum, resulting in atrial septal defects, ventricular septal defects, or damage to valve leaflets or papillary muscles. These internal injuries may be missed at the time of operation, and are often asymptomatic. Progression of injury may also cause focal dyskinesia or cardiac pseudoaneurysm in the postinjury period. Myocardial oedema and haematoma may prevent flow through atrial septal defects or ventricular septal defects, and wall motion abnormalities may take weeks to develop; therefore, it is essential to repeat evaluations a few weeks after injury.

Injuries to the oesophagus

Because of its protected location in the posterior mediastinum, blunt injury to the thoracic oesophagus is exceedingly rare. When it occurs, blunt rupture of the oesophagus is usually seen in the distal third just above the gastro-oesophageal junction, similar to Boerhaave syndrome. The injury is probably produced by sudden increases in the intraoesophageal pressure owing to blunt upper abdominal trauma in the presence of a full stomach.

Penetrating injuries are more common and may occur at any level. One must conduct a thorough search for injury to the oesophagus if a penetrating wound has a trajectory that is in proximity to the thoracic oesophagus. There is a high incidence of major associated injury with either blunt or penetrating mechanisms.

Specific signs and symptoms due to perforation of the oesophagus are few and generally the patient's complaints are related to associated injuries. As noted, early diagnosis is the single most important factor in determining ultimate outcome. Findings on CXR include pneumomediastinum, widened mediastinum, air in the prevertebral space or left pleural effusion. The best initial study is contrast oesophagography, although this test does have a significant false-negative rate; adjunctive flexible oesophagoscopy can add to the sensitivity. Early operative repair is the key to successful therapy. Mediastinal contamination and eventual sepsis lead to prohibitively high mortality rates in patients whose operative repair is delayed beyond 12–24 hours.

The upper two-thirds of the thoracic oesophagus are best approached through a right thoracotomy whereas injuries confined to the lower third may be more easily accessible from the left. The general operative principles are to attempt surgical closure when possible and establish wide mediastinal drainage. Extensive tissue destruction, massive mediastinal contamination or delay in surgery may preclude primary repair and constitute indications for exclusion.

Abdominal and pelvic injuries

The abdomen encompasses a large area of the body from the diaphragm superiorly to the infragluteal fold inferiorly, including the entire circumference of this region. Abdominal injuries are generally classified into blunt or penetrating based on mechanism. This breakdown can be useful to the surgeon because the types of injury, presentation and management are quite different. Approximately 25% of all abdominal trauma patients will require a laparotomy at some point during their hospital stay.

Blunt trauma accounts for >80% of all abdominal trauma, resulting in a rate of intra-abdominal injury of approximately 18%. The incidence of blunt trauma is increasing owing to the increasing rate of automobile and motorcycle accidents. The car remains the cause of the majority of blunt abdominal trauma. Solid organ injury is more common than hollow visceral injury following blunt trauma, with the liver being the most commonly injured organ. Given the success of selective non-operative management of solid organ injury, the need for laparotomy following blunt trauma is low. Blunt abdominal trauma requires exploratory laparotomy in only 4% of cases but accounts for 78% of trauma admissions. Mechanisms of injury in blunt abdominal trauma include direct blunt force against bony surfaces such as the spine or pelvis, acceleration/deceleration injury, 'burst' injury and laceration from broken rib, pelvis or spine fragments.

Penetrating trauma accounts for approximately 20% of trauma admissions. The liver and small bowel are the most commonly injured organs. Surgical exploration is more frequently required following penetrating trauma, with 15% of posterior stab wounds, 50% of anterior stab wounds and 60–75% of gunshot wounds requiring laparotomy. In general, higher velocity mechanisms result in worse outcomes, such that gunshot wounds are worse than stab wounds and military wounds are worse than civilian wounds. While injury following low-velocity penetrating trauma results only from direct laceration, injury following medium- and high-velocity mechanisms can occur as a result of laceration, cavitation or blast. Injuries to both thoracic and abdominal cavities occur in 25% of patients with penetrating trauma as the wound may traverse the diaphragm.

Failure to recognize occult abdominal haemorrhage and to successfully control bleeding leads to significant morbidity, and such injuries account for approximately 10% of traumatic deaths that occur annually in the USA. Additionally, failure to appropriately manage abdominal injuries accounts for the majority of preventable deaths following trauma. Trauma to the abdomen is often associated with multisystem injuries, particularly the central nervous system, chest and musculoskeletal system, and these concomitant injuries may obscure signs and symptoms of abdominal injury. Because of this, the need for repeated assessment of the patient suspected of having intra-abdominal injury cannot be overemphasized. The trauma surgeon must take responsibility for overall treatment and provide consistent serial examinations.

Significant intra-abdominal injury can follow iatrogenic causes, including endoscopy, CPR, peritoneal dialysis, paracentesis, biopsy and contrast enema. The principles of diagnosis and treatment are no different from other traumatic injuries.

Anatomical considerations

A practical knowledge of the contents of the abdomen is important. Abdominal injuries can be divided into intraperitoneal, pelvic and retroperitoneal injuries. Examination of the true abdomen and intraperitoneal structures is relatively straightforward. Blood and enteric contents from both hollow visceral and abdominal vascular injury often cause peritoneal irritation and severe pain prompting operative exploration. In the absence of peritonitis, the FAST examination or DPA/DPL can be useful for detecting intraperitoneal free fluid in the hypotensive patient, thus identifying the abdomen as a source of significant haemorrhage. In contrast, injury to the pelvis and retroperitoneal structures can be a diagnostic challenge. Bleeding into both the pelvis and retroperitoneum is often asymptomatic, even when large volumes have been lost, and is undetectable by FAST as well as by DPA/DPL. Damage to the retroperitoneal portions of hollow viscous structures will likewise often be asymptomatic, or present with vague non-specific findings such as tachycardia or failure to clear a base deficit. Because of this a high level of clinical suspicion and judicious use of adjunctive imaging such as CT scanning, urethrography, and angiography are necessary to avoid delays in diagnosis and treatment.

Diagnosis

A history should be obtained from the patient if possible, and from the paramedics in the MIVT format. Information on the type of weapon involved in assaults and the types of vehicle, location of the patient, presence or absence of restraints and vehicle damage in accidents and even images of the scene can be useful to the clinician.

Upon arrival, ATLS protocols should be followed and blood samples should be drawn for basic studies, most importantly type and cross-match. This is then followed by the abdominal examination as a part of the complete secondary survey. The key objective of the physical examination is to determine whether operative exploration is necessary. Absolute indications for surgery include peritonitis, hypotension with evidence of intraperitoneal haemorrhage and evisceration. Unfortunately, outside of these findings, physical examination can be non-specific. Bleeding may be retroperitoneal, and the physical examination may be benign even in the presence of significant haemorrhage. Alternatively, associated injuries to the abdominal wall, overlying ribs and pelvis may result in significant pain and tenderness on examination even in the absence of any intra-abdominal organ injury.

Abdominal assessment should be systematic and orderly. The abdominal wall should be inspected for seat belt signs, abrasions, ecchymoses, lacerations, wounds and foreign objects. Particular care should be taken if a seat belt sign is found, as the presence of an abdominal seat belt sign is associated with a significant increase in intra-abdominal injuries. In adults the presence of a seat belt sign increases the risk of intra-abdominal injury two- to eightfold. This is further increased in paediatric patients, in whom rates of intra-abdominal injury can be increased as much as 12-fold in the presence of a seat belt sign. Hollow visceral, particularly the duodenum, and mesenteric injuries are markedly increased and the threshold to operate on a patient with free fluid and a seat belt sign should be very low. Pancreatic injury is also increased, particularly among paediatric patients. However, the seat belt sign is not associated with increased mortality.

In all cases, but particularly with penetrating mechanisms, care should be taken to inspect the groins, axillae and back to ensure no additional injuries are missed. The abdomen should be palpated to assess for peritoneal signs or rigidity and localized tenderness. In blunt injuries, the pelvis should be assessed for stability, tenderness and range of motion of the hips. If pelvic fracture is suspected the perineum and urethral meatus should be inspected and a rectal examination performed to rule out distal genitourinary injury. In the absence of signs/symptoms of urethral injury a Foley catheter should be placed to assess for haematuria and resuscitation status. A nasogastric tube should be placed as well to decompress the stomach and evaluate for bleeding. Both gastric and urinary bleeding are indications for laparotomy following penetrating trauma.

The abdominal examination may be supplemented by diagnostic imaging methodologies including FAST, plain radiographs, DPA/DPL and CT scan. Prior to any imaging all penetrating wounds should be marked with radio-opaque clips. Plain radiographs or CT scan can then be taken to delineate the trajectory of the bullet or path of the knife and allow for an intelligent assessment of the likelihood of associated injury. Additionally, wounds can be packed with Betadine-soaked gauze, again in order to delineate the tract of injury.

Choice of imaging is determined by the clinical condition of the patient. Unstable patients with isolated abdominal trauma should proceed directly to the operating theatre; however, those with blunt or multicavitary penetrating trauma may have multiple sources of haemorrhage and it is necessary to identify the correct cavity for exploration prior to transfer to the operating theatre.

FAST remains the most rapid, non-invasive and repeatable modality for determining the presence of free fluid in the abdomen. The purpose of the FAST examination is to identify the presence of free fluid in the peritoneal cavity; however, at least 200–300 mL of blood must accumulate before it can be reliably seen on FAST. While the reported sensitivity of FAST varies from 67% to 80%, the specificity and accuracy are consistently reported to be 98–100%. Therefore, although a positive study is reliable, a negative FAST does not exclude intraperitoneal haemorrhage. A positive FAST in the hypotensive blunt trauma patient warrants immediate exploratory laparotomy. If the FAST is negative or indeterminate in hypotensive patients, DPA or DPL may be of use.

First introduced in 1965, DPL has a sensitivity of 90–95%, specificity of 99–100% and accuracy of 92–98%. Many surgeons perform DPA only and do not add lavage if the aspiration is negative because it is unlikely that blood detected by lavage and not by DPA or FAST is responsible for significant hypotension. The false-positive rate for DPL is higher in patients with pelvic fractures and, similar to FAST, DPL is unreliable in detecting injury to retroperitoneal organs. Patients with a positive DPA, defined as aspiration of ≥10 mL of blood or aspiration of succus, or a grossly positive DPL (drainage of bloody or blood-tinged effluent) should be taken immediately for surgical exploration. In stable patients, if the returning lavage fluid is not frankly positive, microscopic analysis can be performed; a positive result is generally considered as a RBC count >100 000/mm^3, WBC >500/mm^3, amylase >200 units or the presence of bile or faecal material.

When performing a DPA, proper technique is essential to avoid iatrogenic injury and to obtain the best sensitivity. First, a Foley catheter should be placed to decrease risk of bladder injury, a nasogastric tube should be inserted to avoid gastric injury, and the abdomen should be prepared in a sterile manner. The infraumbilical site should be used if there is no evidence of pelvic fracture or pregnancy; if either is present or suspected, the supraumbilical position should be chosen. An incision approximately 1 cm long should be created sharply, then a needle inserted aiming towards the pelvis until a pop is felt as the fascia is breached. A wire is then inserted and the needle removed. Seldinger technique is then used to thread a catheter over the wire into the peritoneal cavity. A syringe is used to aspirate the catheter. If the aspirate is negative and lavage is desired, the syringe is removed leaving the catheter in place, valveless tubing is attached to the catheter and 1 litre of sterile saline for adults or 10 mL/kg for paediatric patients is infused. The tubing should not be disconnected at any point, and once the fluid bolus has infused the tubing should be lowered below the patient to allow gravity to assist in removal of the infusate into a vacuum bottle or empty saline bag. It is important to remember that DPL is only considered adequate if ≥75% of the infusate is returned. DPA/DPL are contraindicated in patients with previous laparotomy, and may be technically difficult in the pregnant or morbidly obese patient.

In the stable patient with evidence of significant abdominal trauma FAST has little value, as a negative scan cannot reliably exclude injury and a positive scan does not necessarily require laparotomy. DPA/DPL may have a role in the stable patient who is unevaluable owing to a depressed level of consciousness, spinal cord injury or intoxication, or in those who will be unevaluable owing to general anaesthesia for surgical treatment of concomitant injury. In stable patients without peritonitis who have a reliable abdominal examination and who are not distracted, intoxicated or anaesthetized, serial abdominal examination has proved a reliable means of excluding surgical abdominal injury, particularly in the patient with abdominal stab wounds. In stable patients with high-velocity penetrating trauma or blunt trauma, serial abdominal examinations should be supplemented with CT scan.

Advantages of CT scan include high sensitivity and specificity – up to 90–100% for solid organ injury; ability to evaluate the entirety of the abdomen, retroperitoneum and pelvis; excellent delineation of tract of injury following penetrating trauma; and high sensitivity for detection of intraperitoneal fluid. CT is also non-invasive, and can be used to reliably predict patients with solid organ and pelvic injury who may benefit from angioembolization. Additionally, when genitourinary injury is suspected because of the location of wounds or haematuria, delayed images can be easily obtained and used to evaluate the urinary collecting system. Disadvantages include cost, radiation and contrast exposure, need to remove the patient from a monitored, therapeutic setting, and poor sensitivity for hollow viscous and diaphragm injury. When present, some indications of hollow viscous injury include bowel wall thickening, mesenteric thickening/haematoma, free fluid without solid organ injury and pneumoperitoneum (Figure 14.23). The presence of any of these findings warrants surgical exploration. However, absence

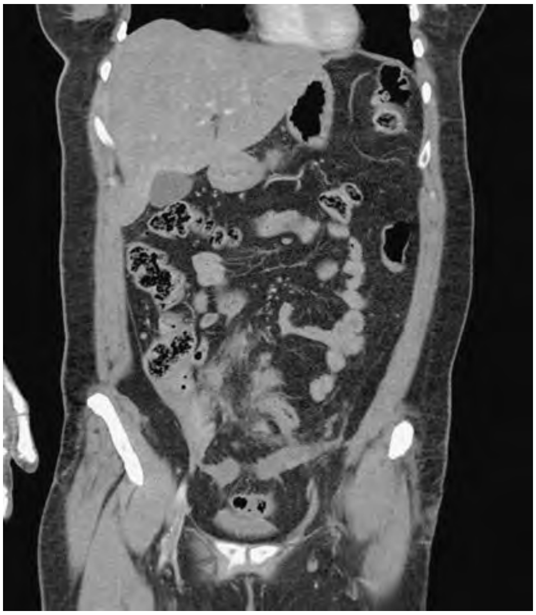


Figure 14.23 Coronal CT revealing free fluid in the pelvis as well as within the mesentery of the small bowel with inflammatory changes of the small bowel mesentery near the right lower quadrant;, these findings are concerning but not diagnostic for hollow viscous injury. At laparotomy the patient was found to have a full thickness mesenteric laceration with active bleeding and an associated antimesenteric small bowel injury.

of these signs does not definitively exclude hollow visceral injury and patients with significant mechanisms of injury require serial abdominal examinations and a high index of suspicion should be maintained by the clinician. Should any deterioration in vital signs, laboratory values or abdominal examination occur, there should be a low threshold for surgical exploration. Patients with negative CT scans and no clinical deterioration can generally be safely discharged after 24–48 hours of serial observation.

Plain films of the abdomen are rarely used, but may be beneficial in cases of impalement to delineate the tract of injury prior to operation and in cases of multiple projectiles, such as shotgun and multiple gunshot wounds, to look for the distribution of foreign bodies (Figure 14.24). Plain films may occasionally reveal pneumoperitoneum (Figure 14.25), which is another indication for surgical exploration.

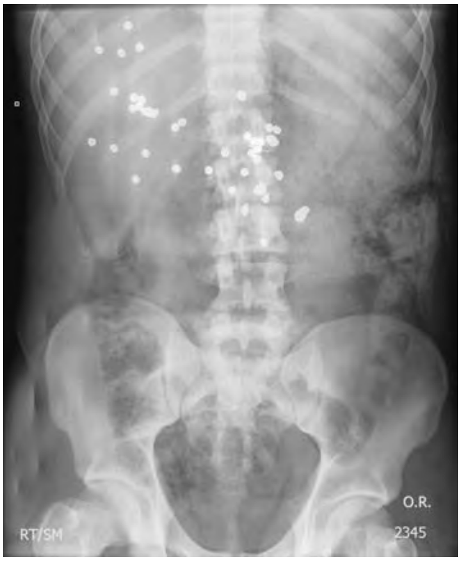


Figure 14.24 Abdominal radiograph revealing the scatter of shotgun pellets in the right upper quadrant and epigastrium.

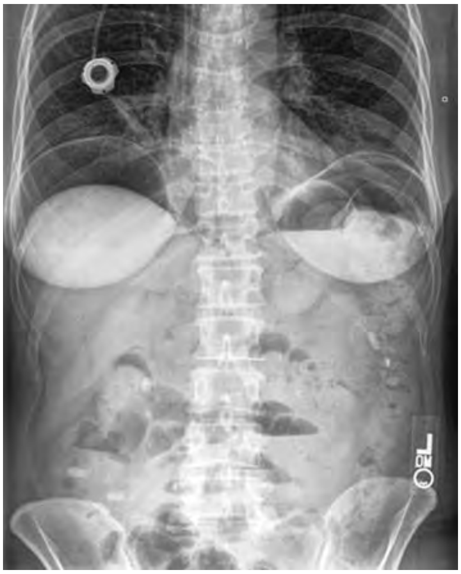


Figure 14.25 Abdominal radiograph revealing a large pneumoperitoneum; note the lucency below the bilateral hemidiaphragms.

Priorities of treatment

It is the job of the trauma surgeon to integrate various specialties that may be called upon to participate in the care of the multiply injured patient. This requires aggressive resuscitation and prioritization of injury work-up and treatment; often, two or more teams may be required to work simultaneously, e.g. in patients with both intra-abdominal and intracranial injury. The trauma surgeon should be aware of what other treatment teams are doing and participate in the decision to alter goals of care from definitive care to damage control for general, neurosurgical and orthopaedic interventions.

In preparation for the operating theatre, particularly in patients with haemodynamic instability, multiple interventions must be co-ordinated. These include large-bore intravenous access with the possible need for central venous catheterization at an appropriate location. Type and cross-match should be performed and MTPs activated appropriately. An arterial cannula should be placed to aid in monitoring resuscitation parameters such as blood pressure, pulse pressure variability, arterial pH and base deficit. Antibiotics should be administered preoperatively, and redosed if massive haemorrhage or long operative times are encountered. Additionally, the trauma surgeon must take aggressive measures to ensure prevention and/or prompt treatment of hypothermia.

Operative approach

The operative approach for abdominal trauma is straightforward. A midline incision is preferred, and there are few reasons to deviate from this. It allows excellent visualization of most intra-abdominal, pelvic and retroperitoneal organs. It also allows extension to a median sternotomy in the event that more proximal control of the vena cava or aorta is needed, or if results of a diaphragmatic window are positive. Because of this the chest should be prepared and included in the sterile field. When preparing the patient the pelvis, including both groins and anterior thighs, should also be included in the sterile field in the event that distal vascular control in the inguinal region is required for pelvic vascular injuries, or if a saphenous vein conduit needs to be harvested for complex vascular reconstructions.

Once the abdomen is opened obvious blood and clot is removed and all four quadrants are packed with sponges. These sponges are then sequentially removed to localize sources of bleeding and/or contamination. Once all areas of active haemorrhage are identified they can be repacked for temporary control. Additional inflow occlusion can be accomplished by clamping the aorta at the diaphragmatic hiatus (Figure 14.26). Obvious sources of intestinal contamination should rapidly be temporarily controlled with clamps, sutures or stapled resection without reconstruction. Retroperitoneal haematomas should be left intact if not ruptured and investigated only once all intra-abdominal catastrophes have been addressed. The major abdominal vessels are all centrally located and, because of this, all central retroperitoneal haematomas should be surgically explored. Lateral and pelvic haematomas are more likely to be caused

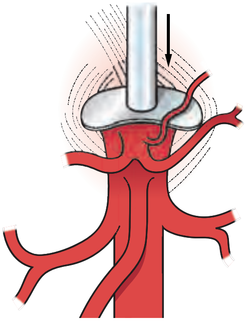


Figure 14.26 The exposed supracoeliac aorta with application of an aortic compression device for inflow occlusion during major abdominal vascular injury. (From Feliciano, http://www.acssurgery.com/acs/chapters/ch0710.htm#.)

by non-surgical bleeding or injury to the kidneys, small unnamed vessels or veins; because of this, lateral and pelvic retroperitoneal haematomas should only be explored if they are expanding, pulsatile or due to penetrating trauma.

Once haemorrhage has been controlled and ongoing contamination stopped, a decision has to be made whether to proceed with definitive repair of all injuries or whether the goals of care have changed to damage control, in which case temporary abdominal closure should be performed.

Damage control surgery was originally created for management of the exsanguinating, severely injured, penetrating trauma patient. Such patients present with extreme physiological derangements, including acidosis due to volume losses, hypoperfusion and anaerobic metabolism; hypothermia; and subsequent coagulopathy (the bloody viscous triad). Prolonged surgery and multicavitary exposure is likely to exacerbate these derangements, contributing to morbidity and mortality. Therefore, the initial surgery should be abbreviated so that treatment of these catastrophic injuries is staged into three distinct phases. First, control of acute haemorrhage and contamination; second, resuscitation; and, third, planned re-exploration for definitive surgical treatment. Indications for damage control surgery include haemodynamic instability, coagulopathy, acidosis (pH <7.2), hypothermia (<35°C), injuries to multiple body cavities, long estimated operative times, and massive transfusion requirements. Injury burden, comorbidities and physiological reserve should also be factored into the decision to pursue damage control surgery. The decision to perform damage control surgery should be made early in the procedure. If damage control surgery has been chosen, once haemorrhage and contamination have been controlled, a temporary abdominal closure should be applied. The temporary abdominal closure must maintain sterility, protect the bowel, prevent evisceration and adhesions, and ideally apply some traction to the abdominal wall to prevent retraction while allowing for expansion of abdominal contents to prevent development of abdominal compartment syndrome. Additionally, the closure should have a means of collecting, removing and quantifying abdominal fluid drainage. Any method that reapproximates fascia or skin should be avoided because of the high risk for abdominal compartment syndrome. The two most popular techniques are vacuum-assisted abdominal dressings and the Bogota bag. Once temporary closure has been achieved, the goal of resuscitation is to reverse acidosis with restoration of circulating blood volume with warm fluids and rewarming with blankets, forced air warmers or ventilation with warmed humidified air. Evidence of return of normal physiology such as normalization of blood pressure, heart rate and temperature; adequate urine output; clearance of elevated lactate levels; and normalization of haemoglobin and coagulation profiles should occur before the patient is returned to the operating theatre for definitive surgical repair and abdominal wall closure. Prior to any definitive abdominal closure care should be taken to carefully inspect the entire gastrointestinal tract from the diaphragmatic hiatus to the peritoneal reflection around the distal rectum, and ensure proper positioning of the nasogastric tube. Additionally, all solid organs should be inspected and palpated, retroperitoneal haematomas assessed for stability, the lesser sac visualized and the diaphragm inspected for any defects.

Specific injuries

Diaphragmatic injuries

Diaphragmatic injury occurs as a result of high-energy acceleration–deceleration trauma, or as a direct laceration from a weapon or broken rib. Blunt injuries tend to be large avulsions, while penetrating injuries tend to be smaller lacerations. Diaphragmatic injury following blunt trauma is rare, occurring in only 3% of cases, in which it is associated with other injuries in 52–100% of cases. Left-sided injuries are more common than right-sided injuries – accounting for 70% of cases. This is probably the result of several factors: first, the posterolateral aspect of the diaphragm is structurally weak on the left; second, the right diaphragm is protected by the liver; and, last, it is more difficult to diagnose injuries on the right. However, right-sided ruptures are associated with more severe injuries and a more significant mechanism. Unlike blunt trauma, diaphragmatic injury may complicate as many as 26% of stab and 13% of gunshot wounds. Injuries are often asymptomatic and imaging continues to have poor sensitivity and specificity. CXR may reveal air-filled organs or NGT above the diaphragm (Figure 14.27a,b), but can also be normal or non-specific in up to 50% of cases. Sensitivity for CT scan (Figure 14.28) and DPL are also poor ranging, from 14% to 61%.

Laparoscopy is the most effective means of both diagnosis and treatment of diaphragmatic injury after penetrating trauma. Sensitivity and negative predictive values are 87.5% and 96.8%, respectively. Any patient with penetrating trauma to the area bounded by the nipples superiorly and costal margin inferiorly should be suspected of having diaphragmatic injury and undergo diagnostic laparoscopy. If injury is found during laparoscopy, the defect can be repaired laparoscopically or the procedure converted to open laparotomy. Pneumothorax can occur with insufflation during laparoscopy and should be decompressed with chest tube placement. If there is another indication for exploratory laparotomy, care should be taken to inspect both diaphragmatic leaflets prior to closure. If the diagnosis is known prior to surgery acute diaphragmatic injury should be approached via the abdomen because of the potential for associated intra-abdominal injury. However, defects discovered remote from the time of injury can be approached through the abdomen, chest or laparoscopically at the discretion of the

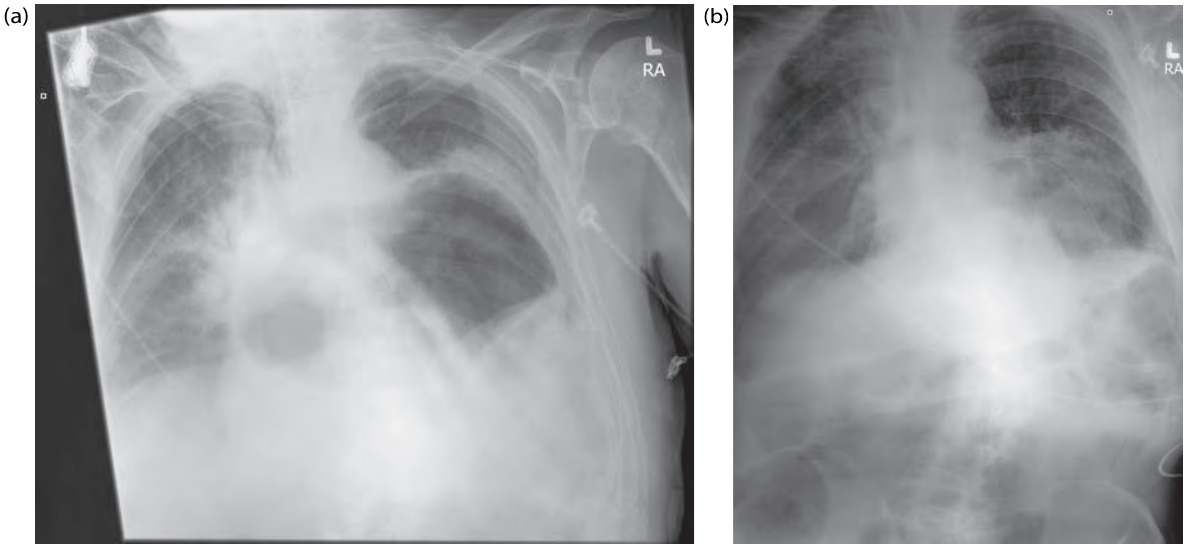


Figure 14.27 (a) Chest radiograph depicting large diaphragm rupture with herniation of the stomach into the left chest before stomach decompression. (b) Same patient as in (a), after placement of a nasogastric tube.

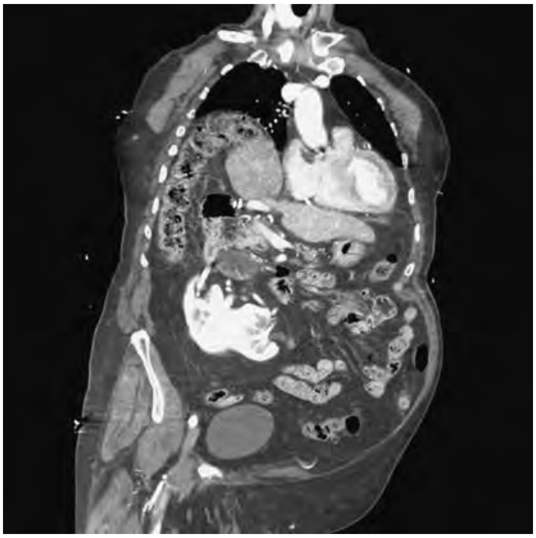


Figure 14.28 Coronal CT scan depicting a right diaphragmatic injury with herniation of the colon and right lobe of the liver into the right chest.

surgeon. All diaphragmatic injuries should be repaired surgically with non-absorbable interrupted sutures. Large injuries, such as those seen with blunt trauma, may require prosthetic material to bridge the defect; Goretex is most often chosen for such repairs. However, care should be taken to avoid placing artificial material into contaminated fields.

Despite careful work-up delay in diagnosis can occur in 30–50% of cases. This delay can range from 7 days to 40 years. During this time injuries can grow as viscera migrate into the thoracic cavity owing to the normal pressure gradient across the diaphragm. Herniation of abdominal contents may cause significant morbidity and mortality as a result of ischaemia of the herniated organs or respiratory distress (Figure 14.28). Mortality associated with diaphragmatic injury ranges from 1% to 28%, primarily related to associated injuries.

Splenic injuries

The spleen is one of the most frequently injured abdominal organs following trauma. The history may include a blow, fall or sports injury to the left flank, lower chest or upper abdomen. Injury to the spleen should also be suspected following penetrating trauma with wounds to the left upper quadrant, flank or back. Splenic injury may be completely asymptomatic, present with vague abdominal or left upper quadrant pain, or referred pain to the left shoulder (Kehr's sign). Stable patients are most likely to be diagnosed by CT scan, which is both highly sensitive and specific and can be used to grade the injury.

Splenic injuries are classified according to the American Association for the Surgery of Trauma Organ Injury Scale (AAST-OIS) (Figure 14.29a,b). Minor injuries include grade I and II, and are minor lacerations or subcapsular haematomas. Moderate injury, i.e. grade III, consists of larger and more severe subcapsular haematomas, lacerations or injuries involving trabecular vessels. Severe injuries include grades IV and V, which are severe lacerations of the segmental or hilar vessels resulting in devascularization or a completely shattered organ.

Management of splenic trauma has undergone a revolution since the 1990s as appreciation of the dangers of abdominal sepsis and overwhelming postsplenectomy sepsis has risen. Non-operative management of solid organ injury, angioembolization and splenic salvage during operative exploration are all acceptable and common alternatives to splenectomy. Rates of success with splenic non-operative management can be as high as 95% for paediatric and 80% for adult populations. Strict contraindications to non-operative management include peritonitis and haemodynamic instability. Studies have now shown acceptable rates of success with non-operative management in patients with neurological injury and in patients with high-grade injury, blush or haemoperitoneum, although these factors increase the risk for failure (Figure 14.30a,b). The addition of angiography with embolization has augmented success rates of non-operative management (Figure 14.31a–d), particularly among patients with

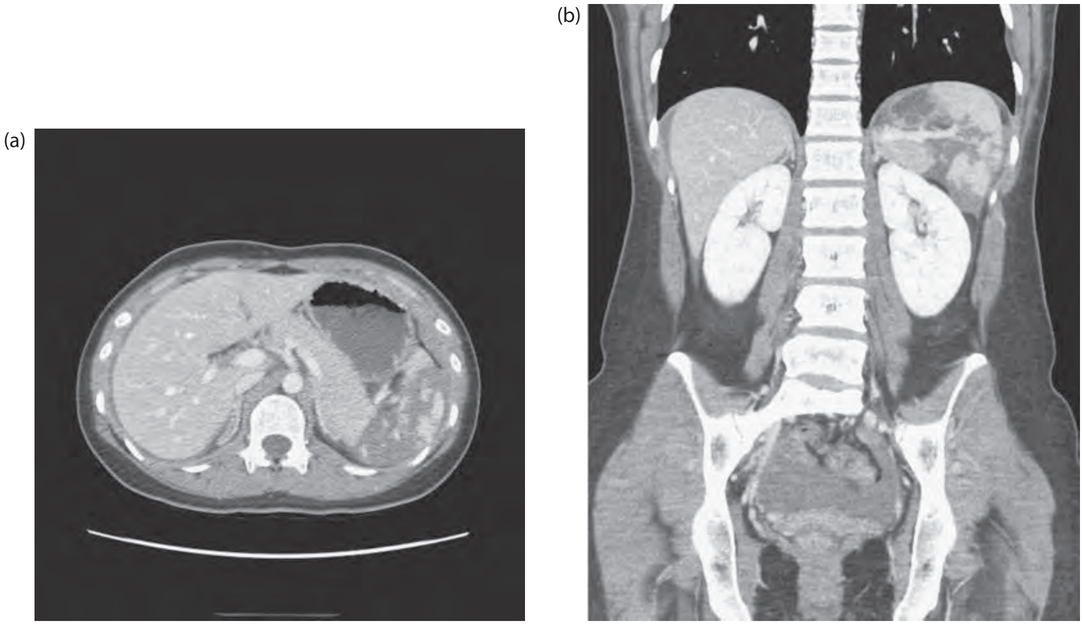


Figure 14.29 (a) Axial and (b) coronal CT scan of a grade IV splenic injury without evidence of active extravasation.

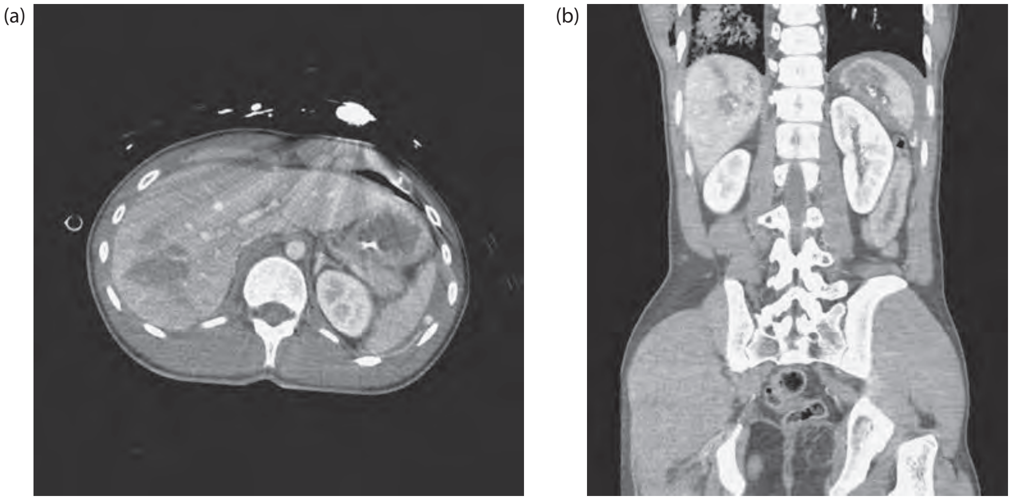


Figure 14.30 (a) Axial and (b) coronal CT scan of a splenic injury with an area of active extravasation in the lateral lower pole. Note that there is also a grade III liver injury to the right lobe.

evidence of arteriovenous fistula, active extravasation and transient responders. Angiography should also strongly be considered in patients with multiple solid organ injuries or concomitant pelvic fracture as it can be both diagnostic and therapeutic in patients with multiple potential sources of haemorrhage.

During operative exploration for another indication, splenic haemorrhage control or failure of non-operative management, the spleen should be carefully evaluated for injury and active bleeding. The process of medially mobilizing the spleen by incising its superior and lateral attachments to the abdominal wall and retroperitoneum is known as the Aird manoeuvre. Care must be taken during mobilization to prevent injury to the diaphragm, stomach and colon. After complete mobilization the spleen and distal pancreas should be delivered into the abdominal incision, allowing visual inspection of the entire organ anteriorly and posteriorly as well as full inspection of the hilum. Bleeding should first be controlled with direct application of digital pressure with a dry sponge. Shallow capsular tears or abraded areas can be controlled with application of topical haemostatic agents. Larger lacerations can be repaired with absorbable sutures. Major injury to the tissues, if localized and <50% of the mass, can be treated with partial splenectomy. However, major vascular injury, injury to >50% of the organ and ongoing bleeding despite splenorrhaphy should be treated with splenectomy (Figure 32). Additionally, splenic salvage techniques should not be pursued among patients with severe physiological derangements, prolonged hypotension or severe concomitant injury, particularly brain injury. With penetrating

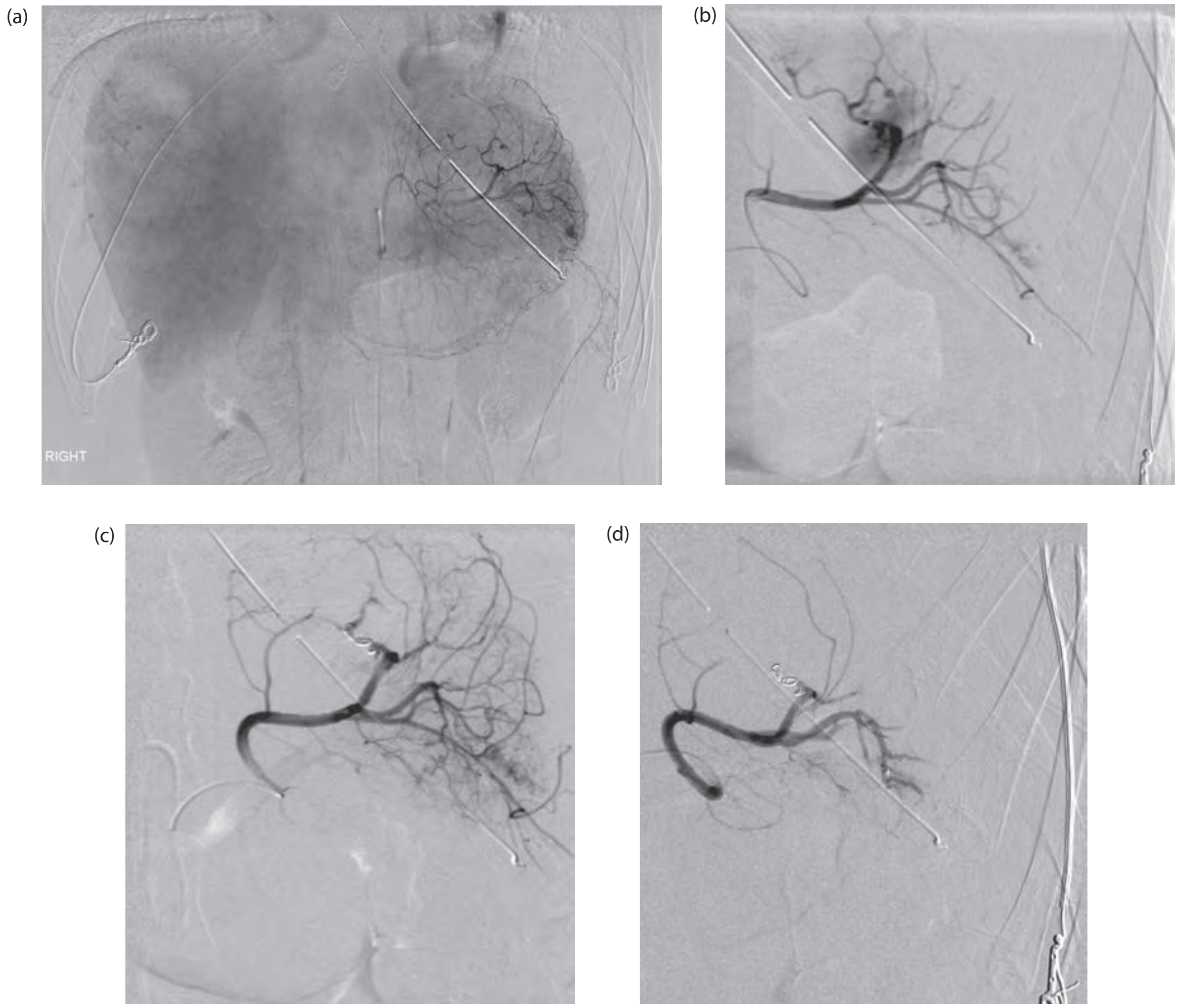


Figure 14.31 (a) Angiogram of the patient in Figure 14.30 showing active extravasation in the lower pole. (b) Selective view of the area of extravasation before coil embolization. (c) Selective view after coil embolization. (d) Selective view after addition of Gelfoam to the coil embolization.

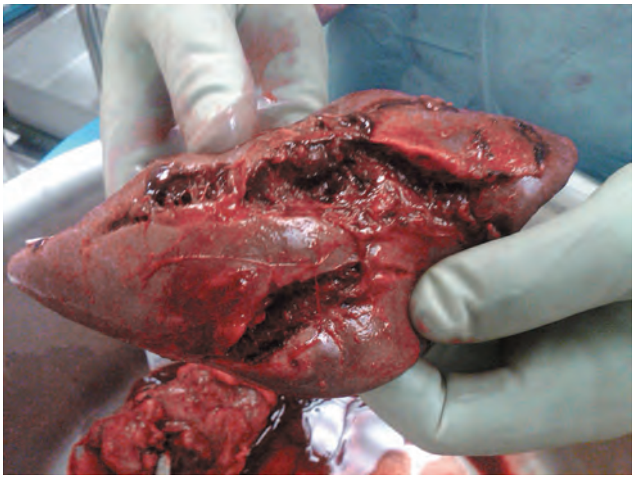


Figure 14.32 An injured spleen after splenectomy.

trauma damage to adjacent structures such as the stomach, colon, pancreas and diaphragm are common and must be investigated and treated.

Complications following splenectomy or splenorrhaphy include early transient thrombocytosis, which does not require treatment and generally resolves in 1–3 months, delayed haemorrhage, pancreatitis, pancreatic leak and abscess.

Postsplenectomy sepsis

The spleen is primarily an immune and haematopoietic organ, producing tuftsin and properdin; postsplenectomy patients are therefore immunosuppressed. Overwhelming postsplenectomy sepsis – the most dreaded and morbid infectious complication – is extremely rare, but may be increased among paediatric patients. However, infectious complications in general, including pneumonia as well as deep and superficial surgical site infections, are significantly increased following traumatic splenectomy even after correcting for injury severity. Postsplenectomy patients are at most risk for infection from encapsulated organisms, such as *Streptococcus pneumoniae, Haemophilus influenzae* and *Neisseria meningitidis*. Following splenectomy patients should receive *Haemophilus,* meningococcal and pneumococcal vaccinations. There is evidence to suggest that vaccinations are most effective if given 14 days following traumatic splenectomy. However, as many patients fail to follow up after trauma, vaccinations should, at a minimum, be given prior to discharge from the hospital. Patients should also receive booster immunizations every 3 years, and be advised of their increased infectious risk. Some clinicians recommend prophylactic antibiotics prior to any invasive dental, ophthalmological and medical procedures as well.

Liver injuries

The liver is the largest organ in the body and injury is a common occurrence following both blunt and penetrating trauma. The overwhelming majority of liver injuries are due to blunt trauma, most commonly motor vehicle accidents. Associated injuries are common following liver injury and occur in up to 80% of cases. Historically the majority of liver injuries were treated with operative exploration and repair. However, because of the low-pressure venous system within normal liver parenchyma, and the tamponade effect of the triangular, coronary and falciform ligaments, over 85% of injuries were found not to be bleeding at the time of surgery. Thus, over time, non-operative management has become the preferred treatment strategy, and the majority of patients tolerate this very well.

Diagnosis

Unfortunately, the history and physical examination are generally insensitive and non-specific for liver injury. The history may reveal blunt force or penetrating wounds to the abdomen, particularly the right upper quadrant or costal margin. Physical findings may be minimal: the abdomen may or may not be distended, tender or have visible findings such as bruising, lacerations, abrasions and abdominal seat belt signs. Diagnosis mainly relies upon imaging and operative findings.

In the hypotensive trauma patient, FAST can be extremely helpful as a positive study identifies the abdomen as the source of bleeding, requiring exploratory laparotomy. Unfortunately, a negative or indeterminate FAST cannot definitively rule out intraperitoneal bleeding. In these cases, DPA should be performed; if aspiration is positive, laparotomy is mandatory.

For stable patients without peritoneal signs CT is the standard method to identify and grade liver injuries according to the AAST-OIS scale for liver injury (Table 14.6). Most injuries are of minor to moderate severity (grades I–III). Some CT findings that have been found to increase the risk of failure of non-operative management include: left lobe injuries, injury into the hepatic veins (Figure 14.33a,b), large haemoperitoneum, laceration >6 cm and porta hepatis involvement. The presence of a 'blush' on CT scan suggests ongoing arterial bleeding or arteriovenous

Table 14.6 Liver injury scale

Grade*	Type of injury	Description of injury
I	Haematoma	Subcapsular, <10% surface area
II	Laceration	Capsular tear, <1 cm parenchymal depth
III	Haematoma	Subcapsular, 10–50% surface area: intraparenchymal <10 cm in diameter
IV	Laceration	Capsular tear 1–3 cm parenchymal depth, <10 cm in length
V	Haematoma	Subcapsular, >50% surface area of ruptured subcapsular or parenchymal haematoma; intraparenchymal haematoma >10 cm or expanding
VI	Laceration	
	Laceration	3 cm parenchymal depth
	Laceration	Parenchymal disruption involving 25–75% hepatic lobe
		Parenchymal disruption involving >75% of hepatic lobe or >3 Couinaud segments within a single lobe
		Juxtahepatic venous injuries, i.e. retrohepatic vena cava/central major hepatic veins
		Hepatic avulsion

*Advance one grade for multiple injuries up to grade IV.
After Moore *et al. J Trauma* 1995;**38**:323.

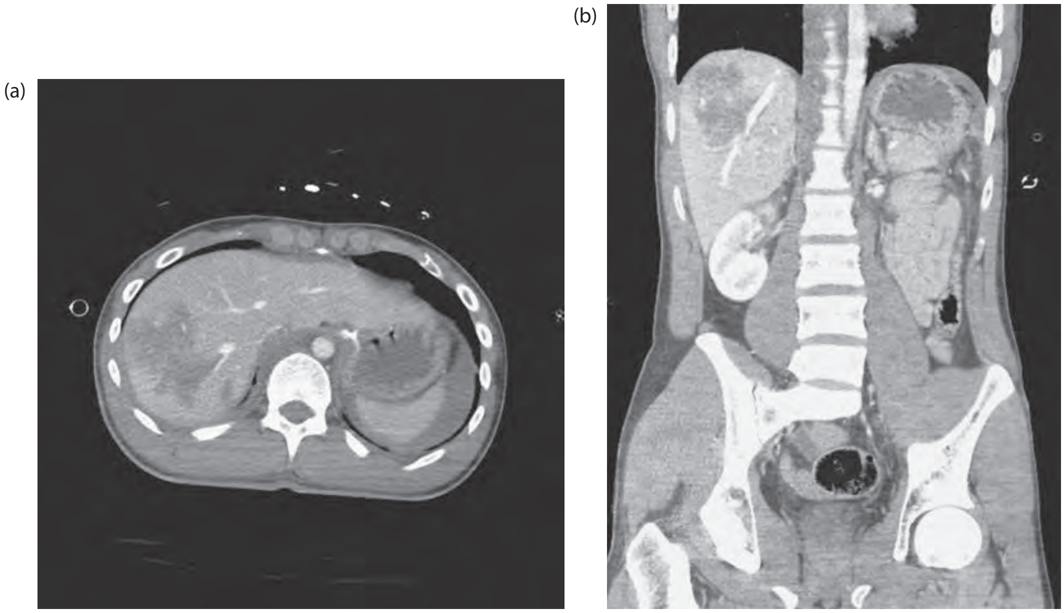


Figure 14.33 (a) Axial and (b) coronal views of grade IV liver injury with involvement of the right hepatic vein.

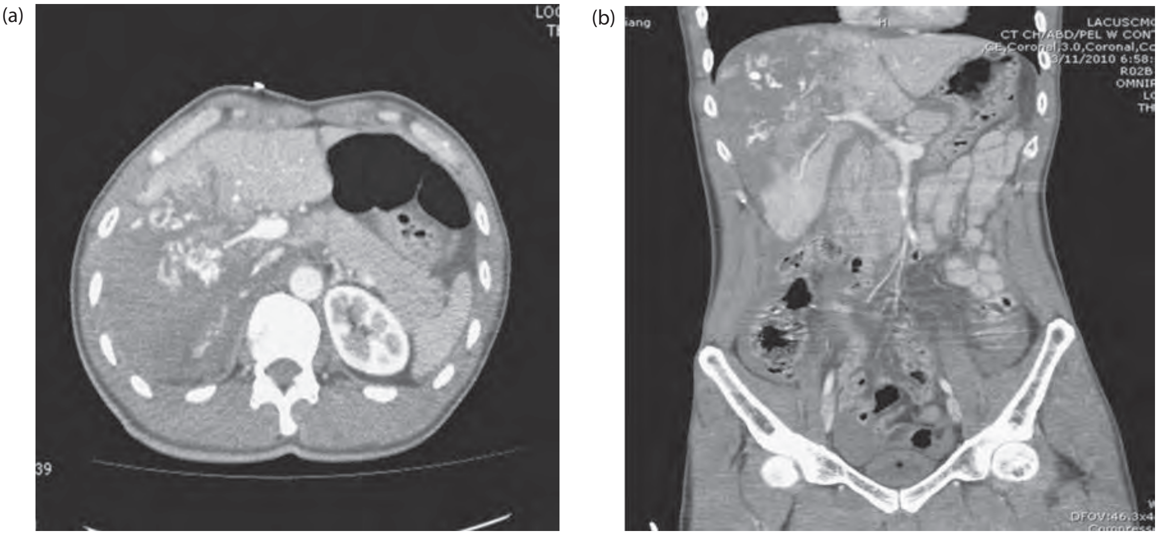


Figure 14.34 (a) Axial and (b) coronal views of a grade V liver injury with involvement of the entire right lobe, extension to the right hepatic and middle hepatic vein as well as areas of active extravasation.

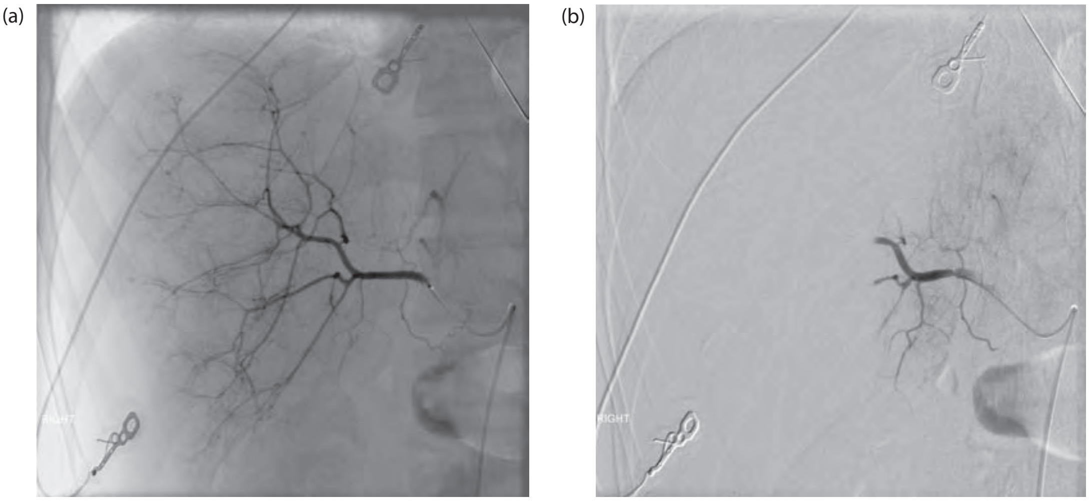


Figure 14.35 Hepatic angiogram (a) before and (b) after embolization of the right hepatic artery.

malformation, and is likely to benefit from angioembolization (Figure 14.34a,b).

Management

Overall success rates for non-operative management of liver injury ranges from 80% to 95%. Early on, non-operative management was used primarily with minor or moderate injuries (grades I–III). However, more recent experience has shown success in up to 40% of stable patients with grade IV and V injuries. The decision to pursue non-operative management should be based not on grade of injury but on clinical findings. Initially, the criteria for expectant management included intact mental status, absence of severe traumatic intracranial bleeding, age <65 years and transfusion of <2 units of PRBCs. Current data indicate that advanced age and comatose states are also no longer strict contraindications to non-operative management.

When performing non-operative management it is important to define a threshold for failure. Clinical predictors of failure include shock on admission, elevated Injury Severity Score, associated solid organ injury, positive FAST and the need for blood transfusion. Failure is generally defined as significant anaemia, transfusion of more than 3–4 units of PRBCs, or the onset of haemodynamic instability or peritonitis. Such failures occur mainly in the first 12–24 hours and rarely occur after 48 hours. When predetermined criteria for failure have been met or appear to be likely, patients should be treated with angiography or surgery as soon as possible.

Stable patients with a blush on CT scan and without peritonitis are candidates for angioembolization (Figure 14.35a,b). Angioembolization can be used before or after surgery and has a significant role in managing late complications such as arteriovenous malformations, pseudoaneurysm and arteriobilious fistula. It is required in 5–6% of patients with liver injury and leads to haemorrhage control in 80–100% of these cases. However, mortality and morbidity can be as high as 27% and 58%, respectively. Complications include hepatic necrosis, abscess and bile leak. A rare complication associated with angioembolization, particularly of the right hepatic artery and its branches, is gallbladder necrosis. Cholecystectomy will be required in these cases.

Angioembolization is an excellent adjunct to non-operative management; however, when there is a precipitous drop in haemoglobin, haemodynamic instability or peritonitis, patients should be taken to the operating theatre, where bleeding is found in 67% of cases. Satisfactory operative repair of significant liver injuries continues to be a challenge even for experienced trauma surgeons. However, the operative principles involved are simple and include: control of bleeding, removal of devitalized tissue and establishment of adequate drainage. In preparing for surgery, consideration should be given to including cell savers, massive transfusion protocols, rapid infusers and adjuncts such as the argon beam.

Once the liver injury has been identified, it should be packed. Extensive mobilization should be avoided in the presence of a haematoma in the ligaments or tissues around the hepatic veins, as this may unleash tamponade of injuries and cause significant bleeding. Packing can be temporary or used for up to 24–72 hours in damage control scenarios. Packing alone results in haemostasis in up to 80% of patients. If packs are to remain in place for a prolonged period of time, an absorbable material such as NuKnit, Surgicell or Vicryl mesh can be interposed between the sponges and the liver parenchyma to prevent bleeding when the packs are removed. Additionally, a dilute solution of saline and hydrogen peroxide can be used to help 'float' the packs off the liver without disturbing underlying clot. However, overpacking should be avoided as it may cause inferior vena cava compression, decreased venous return to the heart and hypotension.

If definitive repair is desired, multiple options are available. Simple lacerations can be controlled with the placement of topical haemostatic agents, argon coagulation or electrocautery. Larger lacerations can be sutured and parenchymal defects filled with viable omentum, creating a tamponade effect and providing macrophages and other cellular mediators of healing. If simple sutures or pressure do not control bleeding from lacerations, the wounds should be opened, providing a tractotomy. This will assist in identification of injured vessels and bile ducts, which can then be individually ligated. If haemorrhage persists, vascular isolation of the liver should be sequentially applied. This starts with clamping of the portal triad within the hepatoduodenal ligament (the Pringle manoeuvre). Although controversy exists, most surgeons agree that the Pringle manoeuvre can be applied safely for at least 30–60 minutes. If a Pringle manoeuvre stops or slows bleeding, the injury is likely to be in a portal venous or hepatic artery tributary. The Pringle manoeuvre can be removed to identify the exact location of injury so that it can be ligated. If selective ligation fails the hepatic artery can be ligated with few consequences as long as the portal flow is not compromised. However, consideration should be given to removal of the gallbladder as ischaemia and necrosis can occur after hepatic artery ligation.

If the Pringle manoeuvre fails to stop the bleeding, haemorrhage from a lacerated hepatic vein or the retrohepatic inferior vena cava is likely. Hepatic vein and juxtahepatic inferior vena cava injuries are rare and associated with very high mortality. Patients with these types of injuries often present in, or near, arrest; in these extreme cases, total vascular occlusion of the liver can be performed to staunch blood flow temporarily and allow the surgeon to identify and directly ligate or repair major vascular injuries. Total vascular isolation requires aortic clamping, a Pringle manoeuvre and occlusion of both the suprarenal and suprahepatic inferior vena cava. An alternative is the atriocaval or Shrock shunt (Figure 14.36). This is a shunt placed in the vena cava either from a venotomy in the infrarenal inferior vena cava or from an incision made in the right atrial appendage. The shunt allows blood flow to return from the renal veins and infrarenal inferior vena cava into the heart, bypassing the suprarenal inferior vena cava and liver. Venovenous bypass is another way to achieve vascular isolation of the liver. Again, a Pringle manoeuvre is performed. Unfortunately, all such advanced vascular techniques are associated with poor outcomes.

Non-anatomic liver resection may be required in up to 3% of cases if there is destructive parenchymal injury (Figure 14.37a,b). In this case, resectional debridement should be expeditiously performed using the finger fracture technique supplemented with the argon beam or electrocautery and selective ligation of encountered vessels. Anatomic liver resection for trauma is rarely required and often has a poor outcome, especially if done for haemorrhage control. This procedure is performed in fewer than 10% of patients, and is associated with 50–67% mortality.

In cases of liver resection, repair of deep or significant lacerations, and extensive or central parenchymal damage, the right upper quadrant should be widely and extensively drained

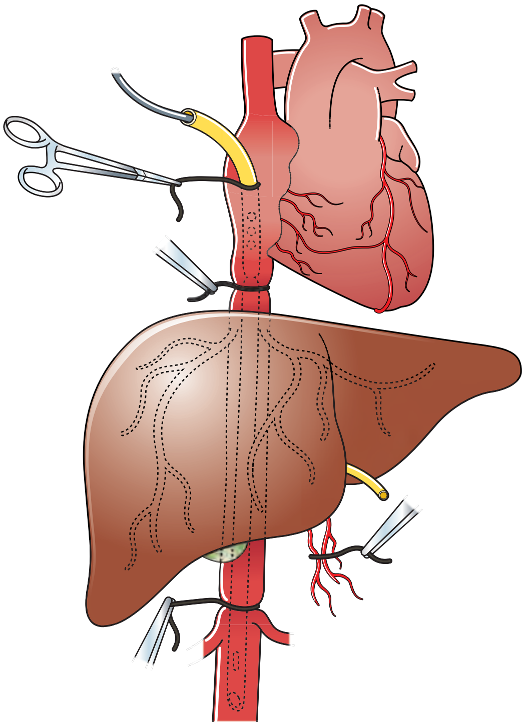


Figure 14.36 The placement of an atriocaval (Shrock) shunt achieved by introducing the catheter into the abdominal vena cava via the right atrial appendage. Sutures or vessel loops above and below the liver achieves proximal and distal control and allows venous return to the heart while completely excluding systemic venous back bleeding from within the liver injury. (From Hoyt *et al.*, *Surg Clin North Am* 2001;**81**:1299–330.)

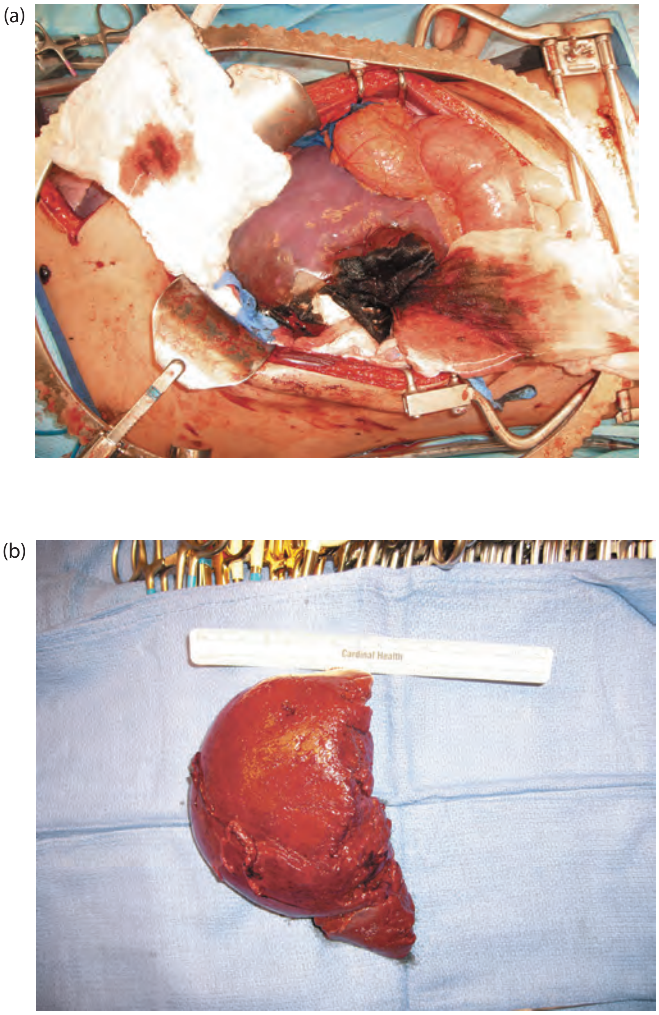


Figure 14.37 (a) A non-anatomical liver resection of the right lobe of the liver following blunt trauma. (b) The resected segment of the right lobe.

with close suction Silastic drains as these injuries are at high risk for developing temporary bile leaks.

Disseminated intravascular coagulation and impaired liver synthesis can occur following liver injury and resection and replacement of coagulation factors and intravascular blood volume are essential during resuscitation. Additionally, replacement and supplementation of magnesium and phosphorus are key during regeneration of liver parenchyma. Patients undergoing major hepatic resection may need continuous infusion of glucose until glycaemic balance can be restored, and aggressive nutritional support should be used.

Complications

The incidence and severity of complications from liver trauma increase with increasing injury grade and if operative management is required. Complications include rebleeding, abscess formation, biloma, biliary fistula and haemobilia. Symptoms of complications include fever, pain, jaundice, leucocytosis, feeding intolerance and tachycardia. Diagnosis of both abscess and biloma is established by ultrasound or CT scan, and both can be treated with percutaneous drainage and antibiotics, with a very high success rate. Persistent bile leaks after adequate drainage should be treated with endoscopic retrograde cholangiopancreatography and sphincterotomy or stent placement. Fistulous connections between the bile ducts and the hepatic vasculature are also possible. A duct-to-artery connection may result in haemobilia and may present with haemoptysis, coffee ground emesis, pain and jaundice. Presentation is often delayed and definitive diagnosis and treatment are established by hepatic angiography and embolization.

Gastric injuries

Stomach

Injuries of the stomach are very rare in blunt trauma but can be common following penetrating trauma. The stomach is partially protected by the rib cage, making gastric injuries relatively difficult to diagnose. Any penetrating wound in the left thoracoabdominal area should be suspected of causing injury to the stomach and requires investigation of the anterior and posterior surface at the time of laparotomy.

A nasogastric tube should be inserted during initial evaluation, and aspiration of blood may point to a gastric injury. The intraoperative evaluation of stomach injury includes good visualization of the oesophageal hiatus and anterior and posterior walls. Adequate evaluation of the posterior wall requires division of the gastrocolic ligament to enter the lesser sac. If there is any question regarding injury, the stomach can be distended with saline and methylene blue to evaluate for leaks. Concomitant injury to the left diaphragm, spleen, transverse colon and splenic flexure should be carefully ruled out during the exploration.

Penetrating wounds are debrided and closed primarily. Maceration of the stomach from significant penetrating or blunt injury may require gastric resection. Postoperative complications include intra-abdominal abscess, particularly in the lesser sac, but these are rare. The other complication of stomach injury is stomach fistula. The treatment is immediate reoperation and repair using healthy tissue. Because of its proximity to the diaphragm, stomach injuries are frequently found in conjunction with diaphragmatic injuries and contamination of the thoracic cavity with gastric contents can occur. In the presence of severe contamination empyema can be a worrisome complication. Drainage of the thoracic cavity with a large-bore chest tube and adequate lavage of the thorax prior to closure of the diaphragmatic injury are important if gastric contents have contaminated the chest.

Duodenum

Isolated injury to the duodenum rarely causes significant hypotension, and signs of peritonitis may be absent or delayed if the injury affects the retroperitoneal portion of the duodenum. Unfortunately, failure to recognize this injury in a timely fashion is associated with high morbidity and mortality caused by abscess formation and sepsis. Entry wounds in the right upper quadrant or between the xiphoid and umbilicus suggest possible injury to the duodenum. Non-penetrating duodenal injury may be caused by crushing as the duodenum is macerated or contused against the spine by a seat belt, steering wheel, handle bar or blunt weapon. Blow-out injuries can also occur as intraperitoneal and extraperitoneal portions of the duodenum create a closed air-filled loop that bursts when compressed (Figure 14.38).

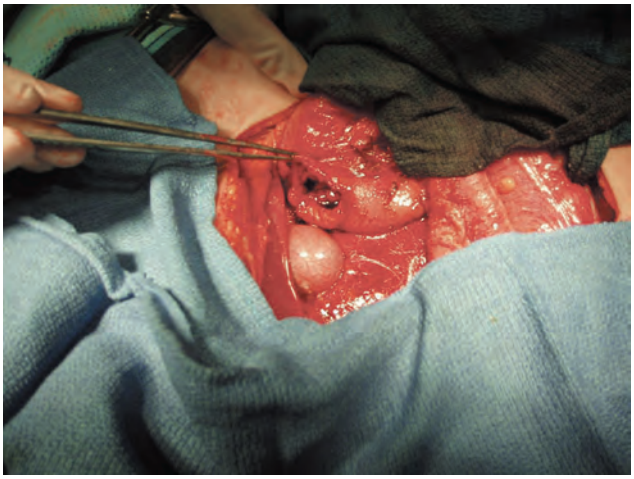


Figure 14.38 A typical antimesenteric blowout injury of the second portion of the duodenum.

Adjunctive diagnostic tests might include hyperamylasaemia. This occurs in about half of the patients with blunt injury to the duodenum as a result of extravasation of intra-abdominal pancreatic amylase. Elevated serum amylase following blunt trauma is not diagnostic of an injury but raises suspicion and necessitates further diagnostic study. Abdominal radiographs may suggest duodenal injury if they show obliteration of the psoas shadow, absence of air in the duodenal bulb or air in the retroperitoneum. Definitive diagnosis requires contrast duodenography or CT scan. Extravasation of contrast material is an absolute indication for laparotomy. Distortion of the duodenum, retroperitoneal air and periduodenal stranding indicate significant injury and are relative indications for surgery. An intramural duodenal haematoma alone without evidence of full thickness perforation is not a definite indication for operative intervention.

Intraoperative evaluation of the duodenum requires complete mobilization of the duodenum (Kocher manoeuvre). The hepatic flexure of the colon is taken down to expose the anterior aspect of the second portion of the duodenum, and inspection of the third and fourth portions of the duodenum at the base of the transverse colon should be done. Mural haematomas should be opened and evacuated to rule out full thickness lacerations that require repair. Retroperitoneal haematomas in the area of the duodenum must be explored and the lesser sac should be entered to exclude associated pancreatic injuries.

Duodenal haematomas in which lacerations have been excluded by contrast radiography or contrast-enhanced CT scans can be managed expectantly with nasogastric decompression until resolution of gastric outlet obstruction. Obstruction can last several weeks and supplemental nutrition via a nasojejunal feeding tube or with parenteral nutrition may be required. In rare cases prolonged obstruction may require surgical decompression; however, this is rare.

Limited perforations or simple lacerations of the duodenum treated within 6 hours of injury are treated with debridement and primary closure. Larger lacerations may require resection and reanastomosis. Suction decompression of the duodenum with a nasogastric or G-tube, as well as external closed suction drainage, is recommended. After 6 hours, the risk of leak increases, and wide drainage and diversion with pyloric exclusion may be required. Multiple or extensive lacerations and complex repairs may also require protection with pyloric exclusion. The pylorus may be sutured via a proximal gastrotomy or externally by application of TA stapler. Gastric drainage is maintained by a gastrojejunostomy. In the majority of cases, the pyloric exclusion, whether sutured or stapled, will open spontaneously a few weeks after surgery; in some instances, if this does not occur endoscopy to dilate a stapled pylorus or cut sutures is often successful in re-establishing an open channel. In extensive injuries and reconstructions thought should also be given to placement of a distal jejunal feeding tube for early enteral access. Historically, extensive duodenal injuries or injuries involving the sphincter of Oddi were decompressed by tube duodenostomy. However, there has been little evidence to suggest that this was effective at decreasing complications or time to healing, and such drains require the creation of additional duodenotomies, which were, themselves, potential sources of leaks and abscesses. As such, traditional tube duodenostomies are no longer advocated, and internal drainage via nasogastric tube and external drainage with Jackson–Pratt drains is now preferred. If additional internal drainage is desired, a jejunal tube can be guided in a retrograde fashion to drain the duodenum.

The distal duodenum (third and fourth portion) can be primarily closed as with the proximal duodenum if the injury is treated early; delays in surgical treatment generally result in poor tissue perfusion and maceration requiring resection and duodenojejunostomy. Pancreaticoduodenectomy (Whipple's procedure for trauma) is occasionally indicated for massive injury or devascularization of the pancreaticoduodenal complex.

Complications of surgical treatment of duodenal injuries include bleeding and leak resulting in duodenal fistulas. Fistulas occur in 5–10% of patients following anastomosis. Unlike a gastric fistula, duodenal fistulas are generally managed non-operatively with nasogastric decompression, nutritional supplementation and aggressive local wound care. Antibiotics are only indicated if there is evidence of infection such as fever, leucocytosis or systemic inflammatory response syndrome. Uncomplicated fistulas will generally resolve in 6 weeks, and operative repair should be considered if they persist beyond this point.

Pancreas

Pancreatic injury following blunt trauma is uncommon, occurring in less than 7% of abdominal trauma cases. Because they tend to have less intraperitoneal and extraperitoneal abdominal fat, children tend to be at increased risk of pancreatic injury. The force required to injure this organ is significant and associated injuries are common, occurring in 70–90% of cases. Anteroposterior compression of the pancreas against the lumbar spine can result in transection at this location adjacent and just to the left of the superior mesenteric vessels. Epigastric and posterior penetrating wounds likewise can penetrate the pancreas and are often associated with significant injuries to the kidney, vena cava and colon. While duct integrity is the main determinant of intervention and outcome, major duct injury is rare, occurring in <15% of pancreatic injuries, and is much more common following penetrating than blunt trauma. Associated injuries are of particular concern as enzymatically active pancreatic juices will increase the risk of anastomotic leaks. The shared blood supply between the pancreas and duodenum makes the likelihood of these two injuries occurring in combination very high. Diagnosis can be quite difficult as clinical findings are likely to be non-specific or non-existent. Suspicion must be raised and serial re-evaluation undertaken if there is any potential for injury. Elevation of serum amylase or lipase following blunt mechanisms is non-specific, but persistent elevation does suggest pancreatic injury and advanced imaging should be performed. CT scanning is most common but is not particularly sensitive, even with multislice scanners and three-dimensional and multiplanar reconstructions. MRI and magnetic resonance cholangiopancreatography may play a role in the diagnosis of pancreatic injury and elucidation of ductal involvement; however, because of the difficulty in obtaining MRI/magnetic resonance cholangiopancreatography particularly after hours, and the lengthy nature of image acquisition, its role may be quite limited.

Patients seldom undergo laparotomy because of pancreatic injury alone. Instead, they are generally operated on because of intraperitoneal blood loss or peritonitis. At the time of laparotomy the pancreas should be inspected and any evidence of adjacent injury excluded. Any retroperitoneal haematoma around the pancreas should be explored; any retroperitoneal bile staining indicates a concurrent duodenal or biliary tract injury which must be repaired. Low-grade injuries require wide surgical drainage and bowel rest only. Closed suction drains should be used as they decrease the rate of septic complications when compared with open or sump drains. Injuries to the distal body and tail that involve the main duct, or are refractory to simple debridement, should be treated with distal pancreatectomy with or without splenectomy. The decision to preserve the spleen should be based primarily on the patient's clinical status: those who are haemodynamically stable, with normal physiological parameters (normothermic, normal coagulation parameters) and few associated injuries may be considered for splenic preservation; however, patients who are unstable, have severe or multiple associated injuries requiring work-up and/or treatment should undergo splenectomy with distal pancreatectomy. If the main duct is injured in a more proximal segment (proximal body, neck), options for management include subtotal pancreatectomy, external drainage with postoperative endoscopic retrograde cholangiopancreatography, and distal drainage with Roux-en-Y pancreaticojejunostomy. If ductal injury is suspected but not definitive, invasive manoeuvres to interrogate the main pancreatic duct such as transection of the tail, duodenotomy and cannulation of the papilla, and cholecystopancreatography should be avoided.

Significant injury to the head of the pancreas or to the right of the superior mesenteric vessels will be associated with a 30–60% probability of temporary pancreatic fistula, and this should be accepted. Severe trauma to the head of the pancreas in association with duodenal injuries should be treated with debridement of the pancreas, closure of the duodenal wound and pyloric exclusion as previously described. Extensive damage to the head of the pancreas and duodenum may require a Whipple procedure. If this is the case the goals of surgery at the time of the initial procedure should be to control blood loss and contamination; reconstruction should be delayed until restoration of circulating blood volume and normal physiology.

The most common complication of pancreatic injury is a persistent pancreatic fistula; if well controlled it should close spontaneously unless there is obstruction to the pancreatic duct. Somatostatin has been used in the treatment of persistent pancreatic fistulas, with some evidence to suggest that its administration may decrease the volume of output; however, there is currently no evidence that somatostatin or octreotide will increase the rate or speed of fistula closure. Abscess, pancreatitis of the pancreatic remnant and pseudocyst may also occur and are generally treated with conservative management consisting of percutaneous drainage, bowel rest and parenteral or distal jejunal nutrition. Antibiotics should be reserved for cases with evidence of active infection. Although rare, if >80% of the pancreatic gland has been resected there is a risk for exocrine insufficiency; this generally can be treated with oral replacement of enzymes. Endocrine insufficiency is an even more rare complication following extensive resection, and is generally treated with alteration of diet and administration of insulin.

Small intestine

Injuries of the small intestine occur in approximately 15–20% of patients who require laparotomy after blunt trauma. The postulated mechanisms of injury include crushing between the spine and a blunt object such as the steering wheel or seat belt, deceleration shearing of the small bowel at fixed points such as the ileocaecal valve and around the superior mesenteric artery, and closed loop rupture caused by increased intra-abdominal pressure (Figure 14.39). Injuries to the small intestine are present in approximately 25–30% of patients who require laparotomy after penetrating trauma, and are due to direct laceration in low-velocity mechanisms, such as stab wounds, and direct laceration, crushing or blast in high-velocity mechanisms.

Diagnosis of hollow visceral injury is primarily clinical. Peritonitis or severe abdominal pain may occur as a result of peritoneal injury or indirectly as a result of irritation from succus or bleeding. Antibiotics should be started preoperatively and redosed if the procedure is prolonged or if there are massive fluid shifts anticipated owing to large blood loss, massive transfusion or a large volume of crystalloid resuscitation. However, prolonged antibiotic administration following completion of the procedure is not recommended, regardless of the amount of contamination encountered. There are no data to suggest an improvement in morbidity or mortality, particularly infectious complications, when antibiotics are continued after 24 hours.

At operation haemostasis is the first priority; when this is adequate and the patient has had a chance to be adequately resuscitated, attention can then be turned to control of contamination. This can be accomplished by application of non-crushing clamps or sutures. The small bowel should be examined from the ligament of Treitz to the ileocaecal valve.

The bowel wall and mesentery should both be examined for defects, lacerations or haematomas. All haematomas affecting the bowel wall should be explored to rule out full thickness injury (Figure 14.40). Large, expanding or pulsatile mesenteric haematomas should also be explored to rule out major vascular injury. Small stable haematomas of the mesentery away from the bowel wall can be observed as long as there is no evidence of ischaemia to the overlying bowel.

Small lacerations can be closed primarily, with care taken to close them transversely in order to prevent stricturing of the repaired bowel segment. Defects due to shotgun, gunshot or blast wounds should be debrided to healthy bleeding tissue prior to repair. It should be kept in mind that injuries due to penetrating mechanisms generally occur in pairs, and a single wound should always be viewed with scepticism and the adjacent bowel carefully examined for missed injury. If two adjacent holes are found, they can be connected across the bridge of bowel and again closed transversely. Large lacerations, macerated tissues or multiple injuries in close proximity benefit from resection and primary anastomosis. Mesenteric defects due to injury or following bowel resection and anastomosis should be closed to prevent internal herniation.

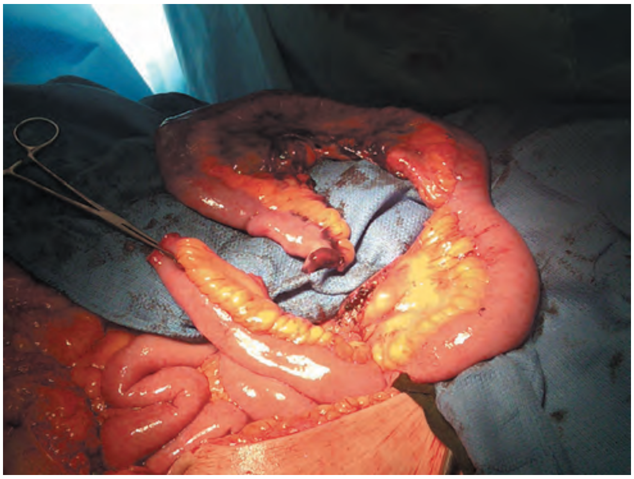


Figure 14.39 An extensive bucket-handle injury of the small bowel with associated complete transection of the bowel wall.

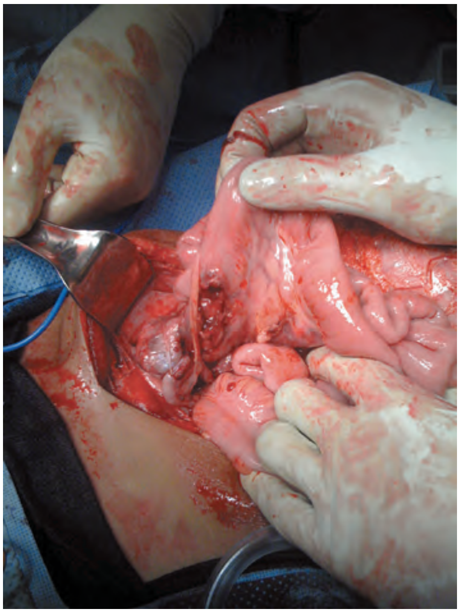


Figure 14.40 A terminal ileal mesenteric haematoma with adjacent full thickness laceration.

Major complications following small bowel injury include abscess, anastomotic leakage, enterocutaneous fistula, wound infection, ileus and anatomical obstruction. Most respond to conservative management with bowel rest and hydration, and antibiotics in cases of active infection and percutaneous drainage of mature fluid collections. Fistulas that are low output are likely to resolve with medical management; however, high-output fistulas, particularly those in the proximal small bowel, may require surgical treatment.

Colon/rectum

The majority of injuries to the colon and rectum are the result of penetrating trauma. The amount of force required to damage the colon is considerable; thus, the colon is relatively refractory to blunt injury, with only 5% of colonic injuries due to blunt mechanisms. Rectal injuries can occur in association with pelvic fracture and any patient with a significant pelvic fracture must have the possibility of rectal injury considered.

Signs and symptoms of colon and rectal injury are non-specific, including peritoneal irritation, leucocytosis and pneumoperitoneum on radiographs. However, patients are often asymptomatic with unremarkable laboratory and radiographic results. Peritoneal lavage is of value if the injury is intraperitoneal and may return fluid with blood or bacteria. If the injury is confined to the extraperitoneal colon or rectum, however, lavage is of no value. The extraperitoneal rectum is particularly vulnerable to injury following penetrating trauma to the buttocks or lower abdomen.

Digital rectal examination is essential. The presence of blood on examination is strong evidence for colon or rectal injury. Proctoscopic and sigmoidoscopic examinations should be performed following penetrating trauma to the buttocks or lower abdomen, or if there is gross blood found on rectal examination with abdominal or pelvic blunt or penetrating trauma. Peritonitis may occur as a result of inflammation of the peritoneum from blood or faeces. Patients with peritonitis, instability, frank blood on digital rectal examination or concerning findings on proctoscopy/sigmoidoscopy should undergo laparotomy. Once the decision to operate has been made, antibiotics should be given prophylactically. If colorectal injury is highly suspected based on physical examination findings, endoscopy or tract of injury, it is essential that broad-spectrum antibiotics with coverage of enteric Gram-negative and anaerobic organisms are chosen; this may include a third-generation cephalosporin with flagyl, or an advanced penicillin such as ampicillin–sulbactam or piperacillin–tazobactam. Use of inappropriate antimicrobial agents such as first-generation cephalosporins has been independently associated with increased risk of postoperative infectious complications. Again, studies have found no additional benefit to continuing antibiotic coverage beyond 24 hours postoperatively.

Simple lacerations of the colon and rectum can be safely repaired primarily after adequate debridement of devitalized tissues. These types of injuries generally heal very well with relatively few complications. In contrast, destructive colon injuries have a very high rate of postoperative complications, ranging from 20% to 40%. Complications include ileus, abscess, wound infection and anastomotic leak. Several factors can significantly increase the rate of complications following surgical repair or resection; these include severe faecal contamination, delay to treatment of >6 hours, transfusion of ≥4 units PRBCs or blood loss >1 litre, hypotension, inappropriate antibiotic choice and more than two associated organ injuries. Options for repair of destructive colon injuries include resection with end-colostomy, resection with primary anastomosis with diversion, and resection with anastomosis but without diversion. The method of repair appears to have no effect on the rate of complications; however, many still consider diversion with or without anastomosis if any of the previously mentioned risk factors for complications are present as the risk of anastomotic leak is high.

Rectal injuries cause morbidity and mortality primarily because of a delay in diagnosis and inadequate initial therapy. Rectal injury must be suspected when there is any penetrating perineal or gluteal injury or a sacral fracture that produces a pelvic ring disruption. Proctosigmoidoscopy is essential for diagnosis and is best done in the operating theatre with the patient in the lithotomy position with both the abdomen and perineum sterilely prepared and draped. Intraperitoneal rectal injuries require laparotomy and repair or resection. Low injuries may be difficult to access and repair and may benefit from resection with end-colostomy in a Hartmann-type procedure.

Treatment of extraperitoneal rectal injuries has evolved over the past decades. Traditionally, these injuries were treated with the triple approach of diversion, presacral drainage and rectal lavage. It now appears that this complex and invasive treatment is unnecessary. Comparisons of patients with and without presacral drainage show no benefit in speed of recovery or prevention of pelvic sepsis from drainage. Findings were similar for rectal lavage, with some studies suggesting worse outcomes with lavage. While transanal repair of the rectal defect may be useful in some patients in whom injuries are easily accessible, there is no role for transperitoneal repair of isolated extraperitoneal rectal injuries as this offers no benefit over diversion alone and often results in increased complication rates. Most studies continue to recommend faecal diversion with either open or laparoscopic loop sigmoid colostomy. Similar to other studies of prophylactic antibiotics, prolonged treatment courses do not decrease rates of infectious complications, and most authors recommend perioperative treatment for only 24 hours.

Complications following rectal injuries include pelvic abscess, rectovesical and rectovaginal fistulas, rectal incontinence and strictures, loss of sexual function and urinary incontinence.

Pelvic injuries

Pelvic fractures have been classified by a multitude of different schema; one of the most useful and widespread systems was created by Young and Burgess (Figure 14.41) and is based on the force vectors involved in creating injury. In this schema fractures are classified into three major groups: lateral compression (LC); anteroposterior compression (AC), traditionally known as open book; and vertical shear (VS), including the Malgaigne fracture. The LC and AC groups are further subdivided into three types (I, II, III) according to increasing injury severity and ligamentous disruption. In general, LC type I and AC type I and II are stable, whereas LC types II and III, AC type III and all VS fractures are unstable.

The diagnosis of pelvic fracture should be suspected in any patient with significant blunt abdominal trauma, pain with palpation of the pelvis, instability of the pelvis with anteroposterior or lateral compression or limited passive range of motion of the hips. Diagnosis is generally confirmed by plain radiography (Figure 14.42), although CT scanning with thin cuts through the bony pelvis allows for a very detailed delineation of fracture, displacement and associated soft-tissue injury (Figure 14.43). The addition of intravenous contrast

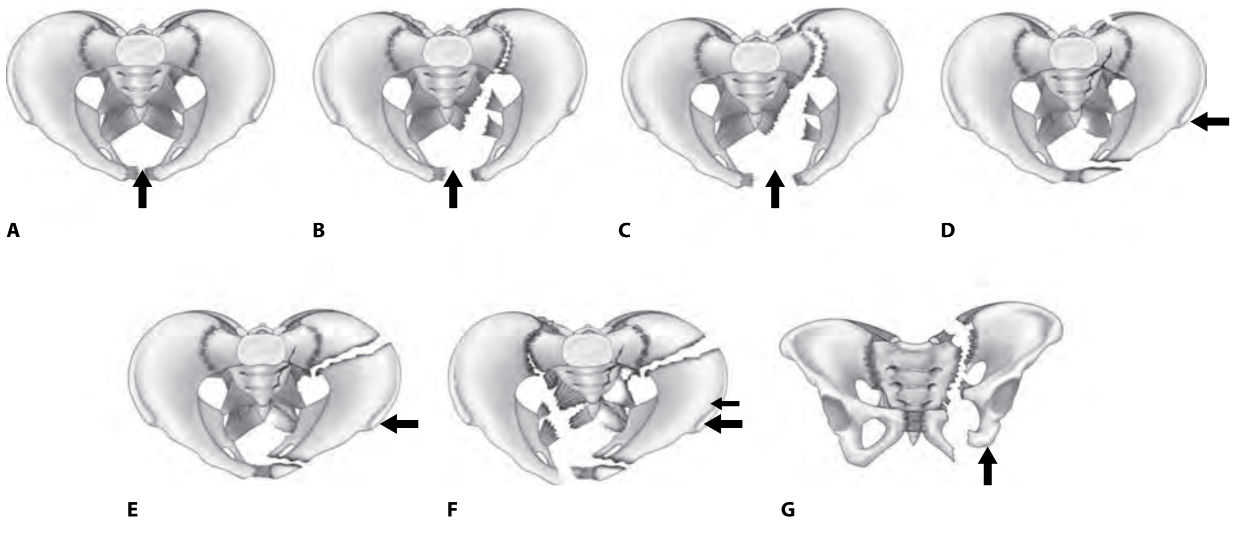


Figure 14.41 Young and Burgess classification of pelvic fractures based on vector of force. (From Hak *et al.*, *J Am Acad Orthop Surg* 2009;**17**:447–57.)

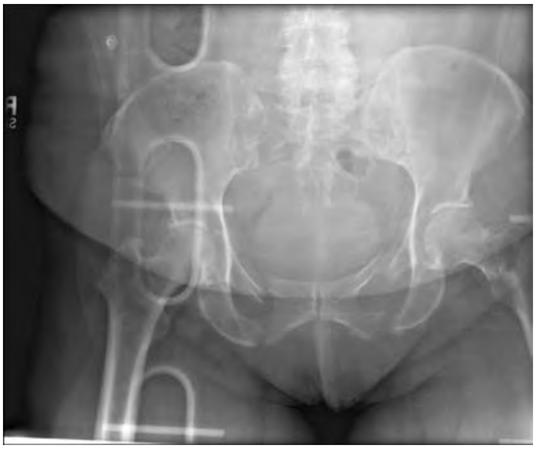


Figure 14.42 Pelvic radiograph; note the bilateral superior and inferior pubic rami fractures consistent with a butterfly fracture-type pattern.

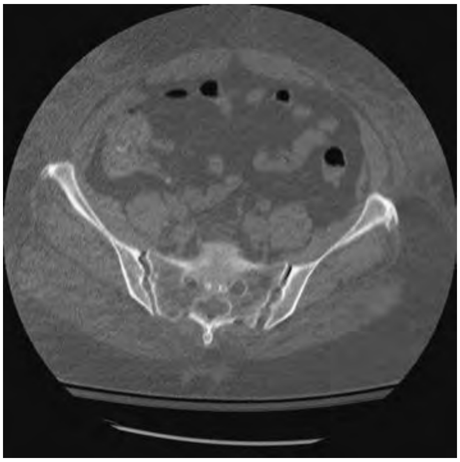


Figure 14.43 Axial CT of the pelvis depicting disruption of the posterior elements and widening of the bilateral sacroiliac joints; note the air within the left sacroiliac joint.

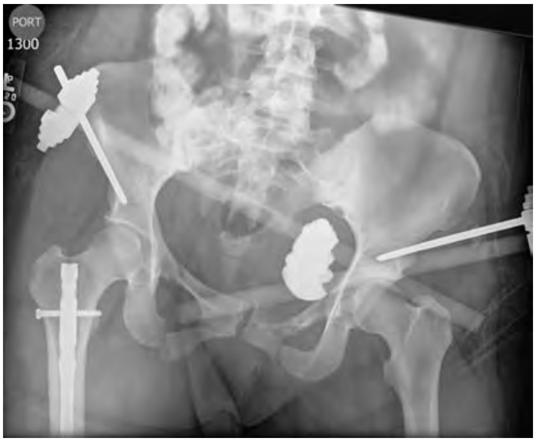


Figure 14.44 Pelvic radiograph demonstrating a butterfly-type fracture after the application of an external fixator.

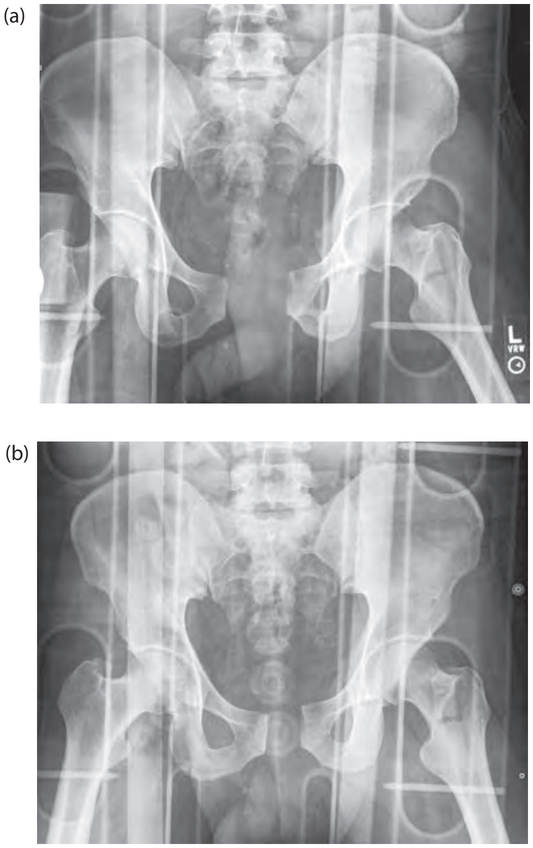


Figure 14.45 Pelvic radiographs depicting an open book fracture with wide pubic symphysis diastasis (a) before external compression and (b) after compression with a sheet.

also allows for detection of active haemorrhage and associated vascular injury; delayed images to assess the ureters and bladder can also be easily obtained with little additional time, and minimal additional contrast exposure. Pelvic fractures are also commonly associated with injuries to the distal genitourinary tract and care should be taken to rule out urethral injury prior to Foley placement and to image the bladder with cystography or CT cystogram in cases of gross haematuria.

Pelvic fractures are a major cause of morbidity and mortality. Blunt trauma is the cause of most injuries, with pedestrian and motor vehicle accidents accounting for the majority. Associated mortality ranges between 10% and 30%, although this can increase to as high as 60% with severe open fractures. Massive haemorrhage and coagulopathy continue to account for 40–60% of the mortality in this group of patients. Bleeding originates from lacerations of both small- and medium-sized arteries and veins of the rich sacral plexus, as well as from fractured bone. Success in managing these patients requires rapid assessment, vigorous resuscitation and reversal of coagulopathy and hypothermia. Haemorrhage can be controlled to some extent by external compression of the pelvis with pelvic binders, external fixators (Figure 14.44) or a bedsheet (Figure 14.45a,b). Binders or sheets are fitted over the anterosuperior iliac spines superiorly, and the femoral heads inferiorly. Stabilization devices close the pelvic ring, decreasing pelvic volume to tamponade bleeding. They also stabilize the broken ends of bone preventing further injury to nearby tissues and decrease pain with repositioning and transport.

In patients who are unstable with free intra-abdominal fluid or suspicion for intra-abdominal injury, laparotomy should be performed. If a stable pelvic haematoma is found it should be left intact, with consideration to postoperative angioembolization.

In expanding or ruptured haematomas with concern for major vascular injury, exploration should be performed. In the absence of identifiable iliac injury consideration can be given to operative bilateral internal iliac artery ligation and embolization with direct injection of a procoagulant such as Gelfoam, thrombin or glue. This performs the same function as angioembolization with a reduction in pelvic vascular inflow. Unstable patients with severe pelvic fracture but without free intraperitoneal fluid should be resuscitated with blood products and application of an external compression device, then taken for angiography. Even among patients who present with haemodynamic instability, angiography is increasingly viewed as an acceptable primary treatment option resulting in haemodynamic stabilization and a survival rate of 72–92%. An alternative or adjunct to angiography in non-responders with negative FAST and DPA is preperitoneal pelvic packing. This method involves making an infraumbilical incision, extending it through the anterior fascia, accessing the preperitoneal space and packing the pelvis. This helps tamponade the pelvic bleeding without decompressing the retroperitoneal haematoma into the peritoneal cavity and has been associated with an excellent ability to achieve haemostasis.

When patients are haemodynamically stable CT has become the diagnostic tool of choice and sensitivity is >90% for detecting pelvic bleeding requiring intervention. The majority of arterial bleeds are due to injuries to branches of the internal iliac artery; injuries to the common and external iliac arteries are rare. In patients with active extravasation on CT, angiography should be considered for diagnostic and therapeutic purposes. Angiography should also be considered for patients without active extravasation on imaging if there is clinical deterioration or transfusion requirements of >4–6 units of PRBCs in the first 24 hours. If significant arterial bleeding is found, selective embolization can be performed. Selective embolization is well tolerated and successful in arresting haemorrhage in 80–100% of cases. Embolization can be performed with Gelfoam, which is delivered easily and is absorbed within a few weeks. In larger vessels and pseudoaneurysms, coils may be required to achieve haemostasis. Gelfoam can be used to supplement coil embolization. In patients in whom selective embolization is not possible unilateral internal iliac artery embolization can be performed with a Gelfoam slurry. If no arterial bleeding is found, bilateral internal iliac artery embolization can also be performed prophylactically. The rich collateral circulation in the pelvis prevents ischaemia in most patients. Major complications following therapeutic or prophylactic angioembolization are rare but include necrosis of pelvic organs or gluteal compartment syndrome. Rebleeding, or continued haemorrhage following therapeutic angioembolization, is also rare; however, many trauma surgeons and interventional radiologists recommend leaving the arterial access sheath in place following the first angiogram for at least 24 hours.

Pelvic fractures can be associated with deep perineal lacerations and involve the rectum or vagina; these injuries are classified as compound pelvic fractures. Compound fractures are associated with a high incidence of septic complications and carry a high mortality – in many cases exceeding 50%. The primary source of contamination is the faecal stream, and diversion with a colostomy is necessary to reduce such complications in most cases.

Urinary tract injuries

General

Because of its location, injuries to the genitourinary tract often are clinically silent and frequently overlooked in the face of more obvious abdominal or thoracic injuries. An awareness of the subtle manifestations of genitourinary injuries is necessary to avoid missing injuries. The physical examination is unreliable in diagnosing urinary tract injuries but they are typically amenable to radiological diagnosis. Systematic, orderly evaluation of the urinary tract reduces the chance of missed injury and limits the number of unnecessary retroperitoneal explorations.

Different criteria should exist for initial evaluation of the urinary tract following blunt and penetrating mechanisms. In blunt trauma, fractures of the lower portion of the rib cage or spinous processes have been associated with an incidence of renal injuries as high as 20%. Findings of a frank haematoma or ecchymosis on physical examination or injury to adjacent solid organs are associated with an increased risk of renal trauma. Penetrating trunk trauma, particularly to the back or flank, has the potential for significant renal injury without any obvious clinical manifestations. A suspicion of renal injury should exist with any penetrating trauma in the vicinity of the renal tract, and in the absence of suspicion for urethral injury a Foley catheter should be inserted. The presence of macroscopic haematuria is indicative of urinary tract injury and should prompt further work-up. However, urinalysis is insensitive and absence of haematuria does not reliably exclude genitourinary injury. If further work-up is necessary the next step is generally CT scan with intravenous contrast and delayed images. Arterial and portal venous phases of the scan are excellent at detecting renal parenchymal injuries, and delayed phases can be obtained to assess the urinary collecting system. Axial CT has the advantage over plain films of providing a 360° view around the bladder, providing excellent sensitivity and specificity for both intra- and extraperitoneal bladder injuries. Additionally CT scan will allow excellent visualization of the surrounding soft tissue and bone, which is particularly helpful in cases of penetrating trauma for diagnosing associated injuries and delineating the tract of the knife or projectile.

If CT scan is not available intravenous pyelography and retrograde cystography can be performed to assess the ureters and bladder. Cystography should be performed with at least 200–300 mL of diluted contrast material. Images should be obtained in more than one plane to decrease the risk of missed injury; this is usually performed in the posteroanterior and oblique orientations. Postvoid images should be obtained in addition to images of the fully distended bladder to improve sensitivity for posterior and extraperitoneal ruptures.

Urethral injury is suspected if penetrating wounds are located in the scrotum, perineum or penile shaft; if, on examination, there is blood at the urethral meatus or a perineal or scrotal haematoma; or if the prostate is absent, free floating or high riding on digital rectal examination. Patients suspected of urethral rupture should not be catheterized until a retrograde urethrogram has been performed. A retrograde urethrogram is performed by inserting a paediatric Foley or red rubber catheter into the very tip of the penis so that it is located in the fossa navicularis; then, the balloon is inflated or a penile clamp is used to occlude the urethral meatus and prevent contrast extravasation. Approximately 30 mL of full-strength contrast is then injected and ideally fluoroscopy is used to assess for extravasation. If fluoroscopy is not available plain radiographs can be taken; again, multiple images in non-congruent planes, primarily lateral and oblique, must be taken. If no contrast extravasation is found a Foley catheter can be placed safely.

Kidney

The kidney is the most commonly injured part of the urinary tract, with injury occurring in 1–3% of all trauma patients and up to 10% of abdominal traumas. Renal injuries are graded according to the AAST-OIS; most injuries are minor, with contusions accounting for the vast majority. Major renal trauma includes deep cortical medullary lacerations involving the collecting system (Figure 14.46) and injuries to the renal vascular pedicle. Blunt mechanisms are far more common than penetrating, accounting for approximately 60% of injuries. Blunt trauma to the renal vessels is more likely to result in thrombosis, whereas penetrating trauma is more likely to bleed (Figure 14.47a,b). Non-operative management is successful in >90% of cases. Risk factors for failure include high-grade injuries, large perinephric haematomas and urinary extravasation. The only absolute contraindication to non-operative management is haemodynamic instability.

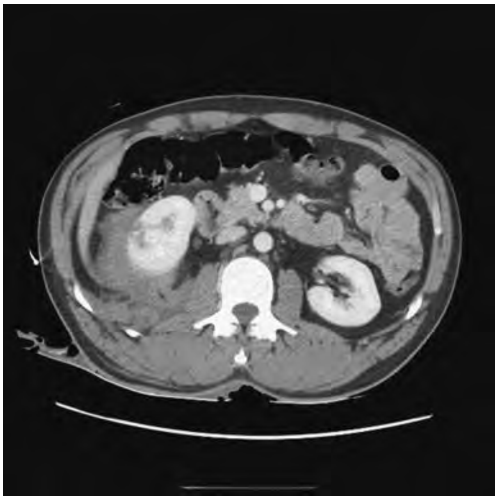


Figure 14.46 Axial CT demonstrating cortical laceration of the right kidney with extension into the urinary pelvis and adjacent perinephric haematoma.

If laparotomy is performed for another injury, and a perinephric or zone 2 haematoma is found, the decision to explore may be difficult. If there is frank haematuria, exploration should be performed to rule out an injury to the collecting system; however, if there is no gross haematuria, and the haematoma is stable, it is reasonable to leave the haematoma as even advanced grade injuries are unlikely to require operative repair, and complications of conservative management generally respond well to percutaneous and endoscopic interventions. If the decision to open the haematoma is made, care should be taken to palpate the contralateral renal fossa to ensure that the opposite kidney is present should rapid nephrectomy be required upon entering the haematoma. The kidney is exposed by medial visceral rotation and can be bluntly delivered from Gerota's fascia. Minor parenchymal injuries can be treated with topical haemostatics. Lacerations should be explored for involvement of the collecting system. If there is no collecting system injury patients can be observed if they are haemostatic or if bleeding can be controlled with simple sutures either with or without pledgets. Absorbable suture should be used to avoid stone formation. Larger parenchymal injuries can be treated with partial nephrectomy. Nephrectomy should be considered based on the degree of injury and the physiological status of the patient.

Complications are more common following nephrorrhaphy than following nephrectomy. Early complications include rebleeding, urine leak and abscess. Late complications include

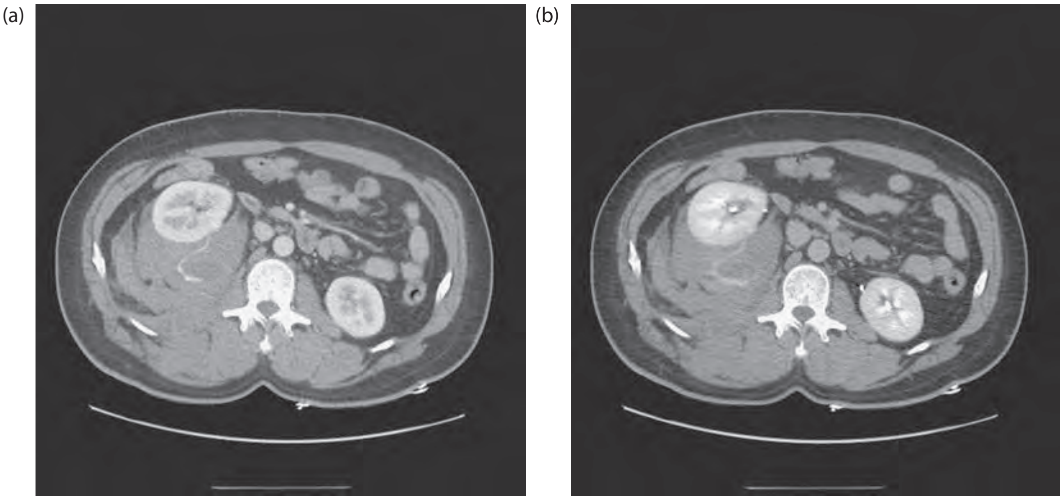


Figure 14.47 Axial CT depicting a right renal cortical laceration with a large adjacent perinephric haematoma and (a) an area of active extravasation during the arterial phase that persists on (b) delayed images consistent with arterial bleeding; note that there is no apparent urinary extravasation from the renal pelvis or proximal ureter.

Page kidney, renovascular hypertension and hydronephrosis. Acute renal dysfunction can occur after traumatic nephrectomy, but tends to be transient and self-limited.

Ureteral injury

Ureteral injury is uncommon and occurs primarily following penetrating trauma. The presence of haematuria is not consistent, as the affected ureter may spasm, particularly following complete transection, and not contribute any blood-tinged urine to the bladder following injury. Ureteral injury should be suspected based on the location of penetrating wounds and the tract of injury seen on CT scan or at the time of laparotomy. Many ureteral injuries are missed at the initial evaluation and present as late urinomas with associated fever, flank mass and pain, or as a urinary fistula with urinary extravasation from the site of the original injury.

In those patients whose clinical condition does not allow for preoperative imaging, the diagnosis of ureteral injury may be made at the time of surgery by the administration of intravenous methylene blue. Extravasation of blue-tinged urine into the operative field serves to localize and confirm the presence of injury.

Surgical options include simple repair, resection with primary anastomosis or reimplantation. All repairs should be made with small absorbable sutures to prevent future complications of stone formation and should be performed over a stent to reduce the risk of stricture.

Bladder

Most bladder injuries occur as a result of blunt external trauma, and should be strongly suspected in patients with haematuria in the presence of pelvic fractures. Bladder rupture may be extraperitoneal or intraperitoneal. The former are usually the result of perforations by adjacent bony fragments from the site of the pelvic fracture. The latter are the result of rupture of the dome that occurs when a full bladder sustains a direct blow. Diagnosis is made by CT or traditional cystography.

Extraperitoneal ruptures are primarily treated conservatively with the placement of a Foley catheter for decompression. The patient is then re-evaluated at regular intervals for resolution of extravasation prior to a voiding trial. Complications can occur in up to 20–25% of patients and include infection, prolonged leak and the need for delayed operative repair. Intraperitoneal injuries require surgical repair through a midline laparotomy. Although previously advocated, placement of a suprapubic catheter is no longer recommended; simple decompression with a Foley catheter in the immediate postoperative period is sufficient.

Urethral

Urethral injuries occur primarily in men following blunt pelvic fracture or straddle injuries, and are uncommon, seen in only 4–10% of patients with pelvic fractures. Blood at the urethral meatus, perineal haematoma, a high-riding prostate on rectal examination, inability to void or gross haematuria are all indicators of urethral injury. When present, suspicion for urethral injury should be high and a retrograde urethrogram should be performed. If there is extravasation of contrast this is diagnostic of urethral disruption. The majority of patients with posterior injuries will have a complete tear, whereas those with anterior injuries will have a complete disruption only half of the time. If the patient is stable an attempt at immediate endoscopic realignment can be made. If a urologist is successful in passing a wire beyond the injury, the Seldinger technique is then used to pass a large Foley catheter over the wire; in essence, this stents the injury open and maintains proper alignment of the proximal and distal urethra, allowing healing to occur while the bladder is adequately drained by the catheter. In unstable patients, or those in whom endoscopic alignment is unsuccessful, a suprapubic catheter should be placed and no further attempts at Foley placement should be made to avoid worsening the severity of the disruption or converting a partial disruption to a complete disruption. Surgical repair of urethral injuries, if necessary, is performed in a delayed fashion. Delayed repair has markedly diminished the incidence of stricture, impotence and incontinence. Concomitant bladder injury is seen in up to 15% of patients with urethral injury, and strong consideration should be given to imaging to rule out this possibility.

Compartment syndrome

Trauma patients with severe intra-abdominal injuries presenting in profound shock and requiring large-volume resuscitation are susceptible to the development of significant bowel wall oedema and sudden increases in intra-abdominal pressure, causing intra-abdominal hypertension. The most accurate way to diagnose intra-abdominal hypertension is with intravesicular measurement. Physical examination has a very poor sensitivity, even in experienced individuals, with sensitivity ranging from 40% to 60%. Intra-abdominal hypertension is defined by the World Society of the Abdominal Compartment Syndrome as intravesicle pressure >12 mmHg. Intra-abdominal hypertension can progress to the point of abdominal compartment syndrome. Abdominal compartment syndrome is defined as sustained intra-abdominal hypertension >20 mmHg associated with new organ dysfunction or failure. Abdominal compartment syndrome can affect many body systems, including neurological, cardiac, pulmonary, gastrointestinal, hepatic and genitourinary (Table 14.7). As with other types of compartment syndrome, the absolute value of intra-abdominal pressure is not as important as the perfusion pressure and the clinical picture. A patient may have abdominal compartment syndrome with organ failure at intra-abdominal pressures <20 mmHg. Increasing use of damage control surgery/temporary abdominal closure may decrease

Table 14.7 Consequences of increased intra-abdominal pressure

Increased	Cardiac rate Pulmonary capillary wedge pressure Peak inspiratory pressure Central venous pressure Intrapleural pressure Systemic vascular resistance
Decreased	Cardiac output Central venous return Visceral blood flow Renal blood flow Glomerular filtration

risks of intra-abdominal hypertension/abdominal compartment syndrome; however, while abdominal compartment syndrome is less likely in the patient with an open abdomen, it is not impossible.

Treatment includes adequate sedation, pain control, pharmacological paralysis, nasogastric decompression, percutaneous catheter decompression and surgical decompression. Non-invasive methods of treatment, especially pharmacological paralysis and catheter decompression, have been shown to be very effective in treating intra-abdominal hypertension, and even abdominal compartment syndrome in certain patient populations.

Vascular trauma

Major vascular injury following trauma is uncommon; however, it can result in extremely high mortality and morbidity. Rapid diagnosis and repair, limiting ischaemic time as much as possible, is crucial to the successful management of these injuries. Modern technological advancements in imaging, prosthetic grafts and endovascular therapies have significantly altered management of vascular injuries. While patients with classic 'hard signs' of vascular injury should still be taken immediately to the operating theatre for open exploration, patients with soft signs of injury now have myriad screening and treatment options. Vascular injuries that were classically difficult to diagnose, expose and definitively treat, such as thoracic aortic and subclavian artery injuries, now have minimally invasive diagnostic and treatment options available as well. One of the most important advancements in the diagnosis of vascular trauma has been the development of CT-A. This has now become the screening tool of choice for traumatic aortic injury, and is becoming more commonly utilized in the diagnosis of upper and lower extremity vascular injury. In the therapeutic realm, minimally invasive endovascular techniques such as stent graft deployment and embolization with coils, glue or Gelfoam are increasingly utilized in the treatment of vascular injury.

Pathophysiology

The spectrum of blood vessel injury consists of transection, laceration, contusion and spasm. Any arterial injury can result in thrombosis, haematoma, pseudoaneurysm or arteriovenous fistula. Complete transection of an artery often induces retraction of the intima and media, which prevents exsanguinating haemorrhage. Partial transection precludes this mechanism and often results in more significant blood loss. Disruption of the intima can lead to thrombosis in the absence of external signs of blood loss or significant haematoma. Traumatic arterial spasm is a rare entity that is difficult to diagnose with certainty, but represents a myogenic reaction independent of the autonomic nervous system in medium-sized muscular arteries. If not recognized early, these injuries can lead to significant morbidity from ischaemia. However, arterial spasm is an angiographic rather than a clinical diagnosis, and it must be emphasized that other types of arterial injury must be ruled out before one can assume that spasm is responsible for poor limb perfusion.

Diagnosis and initial management

Rapid evaluation and restoration of blood flow are essential in order to avoid ischaemic complications of injury. While muscle and peripheral nerve tissue may tolerate anoxia for 4–6 hours, delays in revascularization will lead to intravascular stasis, activation of coagulation and thrombosis. The eventual outcome is irreversible damage to skeletal muscle, peripheral nerves and visceral organ function. Initial management includes airway control, vigorous resuscitation and appropriate prioritization of other accompanying life-threatening injuries. Any external sources of haemorrhage should be controlled with direct application of pressure or tourniquets. Blind probing of a stable wound and attempts at clamping injuries in the emergency department should not be tried as disrupted clot may result in uncontrollable haemorrhage.

Patients with haemodynamic instability, or hard signs of injury, should be transferred immediately to the operating theatre. Patients presenting with soft signs of arterial injury or victims of multilevel penetrating or blunt trauma should undergo angiography for diagnosis and to help determine the optimal surgical approach. Hard signs of vascular injury include arterial bleeding; large, expanding or pulsatile haematoma; hypotension; absence of distal pulses; injury to anatomically related nerves; and thrill or bruit. Soft signs include neurological defect, non-expanding non-pulsatile haematomas and a history of profuse or pulsatile bleeding.

Although vascular injuries resulting in haemorrhage are generally diagnosed promptly, others causing ischaemia or with contained haemorrhage may be more insidious and require a high index of suspicion.

These patients are likely to be haemodynamically stable but may present with signs of ischaemia, including pain, pallor, pulselessness, paralysis and paraesthesiae; however, these generally occur very late in the course of injury. It is paramount to recognize that a palpable pulse or one that can be identified by Doppler imaging does not exclude vascular injury. Pulses may be intact in the presence of injury owing to pulse wave propagation through soft clot, flow past an intimal flap or by way of collateral blood flow around a thrombosed vessel. Any pulse asymmetry, stable haematoma, concerning injury pattern or significant extremity fracture/dislocation should prompt further work-up.

In the extremities, the primary screening examinations are the ABI or BBI. The blood pressure, taken below the level of injury in the affected extremity, is compared with the blood pressure in a non-affected upper extremity. The ABI/BBI has a very good sensitivity, ranging from 95% to 100%, a negative predictive value of 98% and can be easily and rapidly performed at the bedside. If the ABI/BBI is <0.9 further imaging is mandatory as up to 35% of these patients will have a surgically significant injury.

If the ABI or BBI is abnormal, or there are soft signs of injury, further imaging should be performed to definitively diagnose or rule out vascular injury. Options include ultrasonography, CT-A and traditional angiography. Each test has its own set of advantages and disadvantages. The choice of imaging study should primarily depend on institutional availability, reliability and patient presentation. Ultrasonography has the benefits of portability and can be performed in the operating theatre in cases of instability, it is non-invasive and does not require contrast administration. However, it is operator dependent, and has poor sensitivity for thoracoabdominal injuries and in the presence of severe soft-tissue injury. CT-A is an excellent choice for the stable patient, and has very high sensitivity and specificity for cervical, thoracoabdominal and proximal extremity vascular injury. However, it is not appropriate for haemodynamically unstable patients, and requires contrast administration. Angiography has been the traditional gold standard for diagnosis of both truncal and extremity vascular injury, is excellent for preoperative planning and can be both diagnostic and therapeutic. However, it may not be immediately available, requires specialized equipment and personnel, exposes the patient to intravenous contrast and ionizing radiation, and requires arterial puncture.

General operative principles

When preparing for surgery endotracheal intubation must be carefully performed and can be hazardous, particularly when a patient has vascular injuries of the neck and great vessels. The patient should undergo skin preparation and be draped widely with consideration given to potential sources of autogenous vein for subsequent repair prior to induction of anaesthesia. Broad-spectrum antibiotics are routinely administered prior to surgery and in the immediate postoperative period. MTPs should be instituted and adjuncts such as cell saver or other blood reclamation techniques mobilized. Every effort should be made to maintain normothermia as a core temperature of <34°C has been found to be a significant predictor of mortality among these patients.

Incisions should generally be made parallel to the injured vessel with transverse extensions at joint creases. Attempts should be made to preserve arterial and venous collaterals. The time-honoured principle of gaining proximal and distal control prior to evacuation of haematomas remains critical. In situations in which proximal or distal access is difficult, the use of balloon catheters for occlusion has been found to be invaluable. Systemic heparin at a dose of 50–75 units/kg can be administered prior to clamping in isolated vascular injuries; however, multisystem trauma often contraindicates its use. In contrast, regional heparinization is appropriate for most cases that prohibit systemic anticoagulation. Careful use of balloon catheter thrombectomy is essential prior to vascular repair to remove proximal and distal thrombi, minimizing postoperative thromboembolic complications.

Arterial repair can often be performed by simple lateral arteriorrhaphy or resection with primary anastomosis or interposition graft with either autogenous saphenous vein or prosthetic grafts. It is important to adequately debride damaged arterial tissue. Experimental and clinical data indicate that polytetrafluoroethylene (PTFE) is the most infection resistant of available prosthetic materials. Vascular injuries in the presence of massive contamination may be treated by placing an extra-anatomic bypass to provide flow continuity; however, some studies have reported no significant increase in the incidence of graft infection even when placed in a grossly contaminated site.

Accompanying venous injuries, particularly to the deep femoral and popliteal veins, should be repaired to maximize the chances of success when the artery is repaired. Adjunctive intermittent venous compression stockings and elevation of the injured limb should be utilized to prevent oedema formation and development of postphlebitic syndrome. It should also be kept in mind that the risk of compartment syndrome increases significantly in the presence of concomitant arterial and venous injury.

Management of specific injuries

Intrathoracic aorta and great vessels

Thoracoabdominal vascular trauma is a rare entity, accounting for 0.01–2.00% of all trauma admissions. Penetrating mechanisms are responsible for the vast majority of injuries; of blunt causes, motor vehicle crashes are the most common, followed by falls from height and pedestrians struck by vehicles. The morbidity and mortality rates of these injuries are among the highest in vascular trauma. Most of these patients expire prior to arrival at the hospital from profound shock secondary to massive haemorrhage. Successful management of these injuries depends upon aggressive resuscitation, early intubation, prompt diagnosis and rapid surgical control of haemorrhage.

Unlike extremity vascular injury in which external evidence of injury is common, besides hypotension, there are few clinical signs of truncal vascular injury. Some diagnostic signs suggestive of great vessel trauma are described in Box 14.1. Radiographic adjuncts can be very useful in diagnosis of truncal vascular injury but should not delay surgery in unstable patients. Once in the operating theatre, MTPs should be activated and rapid infusers and blood reclamation tools should be utilized. Wide preparation of the neck and chest is essential as the aorta and its branches can be difficult to access and expose, and a variety of incisions or combinations may be required (Figure 14.48).

BOX 14.1 Signs suggestive of major thoracic vascular injuries

- Cardiac arrest
- Persistent shock
- Cardiac tamponade
- A mediastinum widened to >8 cm
- Recurring haemothorax
- Blunting of the aortic knob
- Pleural capping
- Deviation of the left mainstem bronchus
- Displacement of the nasogastric tube
- Neurological deficits
- Pulse deficits
- Bruits

However, about one-third of these injuries will have no obvious clinical signs of vascular trauma, except for a penetrating cutaneous wound. High-grade arteriography is useful, but should not delay surgery in the haemodynamically unstable patient.

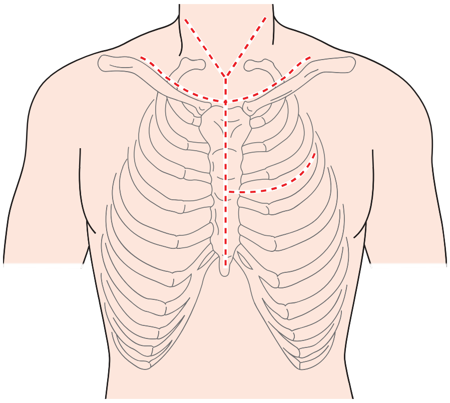


Figure 14.48 Surgical approaches for thoracic aorta and proximal great vessel injuries. (From Hoyt *et al*., *Surg Clin North Am* 2001;**81**:1299–330.)

Innominate vascular injury

Access to the innominate vessels is best accomplished through a median sternotomy. Although minor injuries can be repaired primarily or with a patch graft, major injuries require bypass grafting, usually originating from the proximal aorta. Prosthetic grafts are usually used, although the saphenous vein is an acceptable alternative. If there is significant contamination, ligation and extrathoracic bypass (carotid–carotid, subclavian-subclavian, axillary–axillary, subclavian-carotid) through uncontaminated tissue planes has been used successfully. During repair of the innominate artery, one must consider the need for adequate cerebral blood flow and utilization of an intraluminal shunt. Poor distal back bleeding, stump pressures in the carotid artery <60 mmHg or electroencephalographic changes during intraoperative monitoring suggest the need for shunt placement to maintain cerebral perfusion while repairing the innominate artery.

Accompanying injuries to the superior vena cava and innominate vein are common. The superior vena cava should always be repaired and may require temporary shunting to allow venous return. Although repair of innominate venous injuries is desirable, unilateral ligation is tolerated in most cases. Bilateral ligation results in a superior vena cava-like syndrome. Both superior vena cava and innominate vein injuries may be a source for air embolization. In the event of this complication, immediate direct aspiration of air from the pulmonary artery, right ventricle, atrium and superior vena cava is required.

Subclavian vascular injuries

Subclavian vessel injuries are primarily due to penetrating mechanisms, but are rare, affecting <3% of all penetrating traumas. Injuries associated with blunt trauma are even more uncommon. Injuries often also have accompanying spinal cord or brachial plexus injuries, which are responsible for significant and sometimes devastating disability. These injuries are highly lethal, and patients often present with hypotension, hard signs of vascular injury or compartment syndrome, and are taken for emergency surgery. In rare instances, patients present in stable condition with soft signs of injury, or are asymptomatic; in these cases, CT-A can be very helpful in detecting injuries, for preoperative and endovascular planning. Several large series have found excellent sensitivity and specificity – near 95–100%. Traditional angiography may also be utilized and may be both diagnostic and therapeutic.

Immediate operative repair has been the mainstay of treatment for subclavian arterial and venous injuries. Exposure of the vessels can be performed through a periclavicular or trap-door incision. More proximal injuries can be accessed by median sternotomy for the right, and a high lateral thoracotomy on the left. Because of the posterior location and intimate relationship of the subclavian artery to the brachial plexus, operative exposure and repair can be incredibly difficult. Once localized primary repair is preferred, however, tissue loss often makes this unfeasible as this vessel has very little mobility; in these cases, the artery may be ligated or shunted, venous injuries may be ligated with relative impunity. With any operative repair there is significant risk of bleeding, nerve injury and, in the case of ligation, compartment syndrome and ischaemia. Because of this, alternatives to open surgical repair are very attractive, and there is growing experience with covered stents for treatment of subclavian artery injuries. Several case series have reported mortality rates ranging from 0% to 33%, with technical success rates ranging from 67% to 100% and primary patency rates of 83–100%. These series are, however, limited by relatively short-term follow-up and small study populations.

Patients who are unstable should not considered be for endovascular therapy; other contraindications to endovascular repair include vessel transection and a lack of a proximal vascular fixation site. In some cases stent–graft deployment may occlude the origins of one of the subclavian artery branches, the most important of which are the vertebral and internal mammary arteries. If the vertebral artery may be occluded it is imperative to assess whether the contralateral vertebral circulation is intact, and whether cerebral perfusion will be maintained after occlusion. The internal mammary may be sacrificed, however, at the cost of future use for cardiac revascularization.

Carotid artery injuries

Penetrating trauma accounts for the majority of carotid injuries, and the common carotid is the most frequently injured segment. Carotid injury is often accompanied by injury to vascular, nervous or aerodigestive injury. Injuries due to penetrating trauma most often present with haemorrhage, haematoma, a threatened airway or shock. Associated mortality remains high, and persistent neurological deficit is common.

Carotid injuries are most often approached through a vertical incision along the anterior border of the sternocleidomastoid muscle. Adequate debridement or resection followed by primary anastomosis or interposition grafting is utilized to repair the defect. Shunts should be liberally used when there is concern regarding the adequacy of collateral perfusion. A quick measure of collateral perfusion is the stump pressure: stump pressures <40–55 mmHg are considered by most to be an indication for shunt use. However, few studies have found any significant benefit among patients in whom shunts were used. Reconstruction is associated with better functional outcomes than ligation; therefore, ligation should be reserved for cases of severe physiological derangement, severe concomitant brain injury, a documented ischaemic lesion in the distribution of the injured vessel, in the presence of severe distal thrombosis and with very high skull base injuries.

There has been concern that restoration of cerebral blood flow in the presence of a severe neurological deficit may convert ischaemic infarction to haemorrhagic infarction, and a high mortality rate has been associated with operative repair among these patients. However, neurological impairment may be due to hypoxia and shock associated with carotid injury, rather than directly resulting from cerebral ischaemia, and may reverse with resuscitation and intubation. There are also data to suggest that an aggressive approach to revascularization even in the setting of a significant deficit may result in neurological improvement. Most authors now favour carotid reconstruction in patients with mild to moderate deficits when antegrade flow has been preserved.

Occasionally patients with penetrating carotid injuries may present with haemodynamic stability. In these cases, if there are no hard signs of vascular injury CT-A has become very useful. In these stable patients consideration can also be given to endovascular therapies. These may be particularly attractive in zone 1 and zone 3 injuries, which are technically difficult to access and repair. Small series have found good technical success and low associated mortality with both covered and bare stents as well as coils in the treatment of pseudoaneurysm, arteriovenous fistula and dissection. However, there are some data to suggest increased morbidity due to stent placement, particularly following blunt carotid injuries.

Vertebral artery injuries

The vertebral arteries originate from the subclavian arteries bilaterally and enter the neck in zone 3 laterally. They enter the transverse foramen at the level of C6, and superior to this point they are contained within the bony vertebral canal before entering the base of the skull through the foramen magnum. Traumatic injury of the vertebral artery is rare because of this bony protection. Penetrating trauma accounts for the majority of injuries. CT-A is an excellent screening test in patients with suspect wounds, or concerning blunt mechanisms (hyperextension/rotation). Definite injuries should undergo angiography as they can result in significant haemorrhage, thrombosis or development of a pseudoaneurysm or arteriovenous fistula. Because of the confluence of the vertebral arteries to form the basilar artery, ligation alone is inadequate treatment for most injuries. Percutaneous occlusion with detachable balloons is a useful adjunct in treating these injuries.

Intra-abdominal vascular injuries

Injuries to major intra-abdominal vascular structures are rare and associated with both significant haemorrhage and risk of visceral ischaemia. Penetrating mechanisms are responsible for the vast majority of vascular injuries in the abdomen. Of blunt causes motor vehicle crashes are the most common. Although optimal treatment is repair and reconstruction of normal vascular pathways, knowledge of the anatomy and collateral circulation allows ligation of many injuries when necessary.

Presentation is variable. If the injury has been contained in the retroperitoneum the patient may be relatively stable; however, with free rupture into the peritoneum, the patient will probably present *in extremis*.

Most patients will be in an advanced stage of haemorrhagic shock and immediate large-bore intravenous access, resuscitation and rapid surgical control of bleeding are essential. MTPs should be instituted as soon as abdominal vascular injury is suspected. Despite modern advancements in surgery, mortality from abdominal vascular injuries remains high and increases with increasing number of vessels injured and in the presence of associated non-vascular injuries. In general, venous injuries are the more poorly tolerated and result in a mortality of 50% in many series. Additionally, these injuries pose a technical challenge in exposure and repair or reconstruction of the thin venous walls, which are easily torn. Because flow rates are low, venous repairs are more prone to thrombosis, particularly when prosthetic material is used. When venous ligation is required, fluid sequestration occurs in the affected vascular bed and patients require additional fluid administration to maintain intravascular volume until collateral flow improves. This third spacing significantly increases risks of abdominal compartment syndrome and organ dysfunction. When collateral flow is inadequate, infarction secondary to venous obstruction can occur (Figure 14.49).

The majority of vascular structures are located in the retroperitoneum, which is divided into three zones (Figure 14.50). Zone 1 spans the midline of the abdomen and can be divided into the supramesocolic and inframesocolic regions; it contains the aorta, inferior vena cava, and the coeliac, superior and inferior mesenteric arteries. Zone 2 is located in the paracolic gutters bilaterally and contains the renal vessels and kidneys. Zone 3 begins at the sacral promontory and contains the iliac vessels. The remaining major abdominal vessels, the portal and superior mesenteric veins, are intraperitoneal. All zone 1 haematomas should be explored owing to the high risk of major vascular, pancreatic and duodenal injury. Zone 2

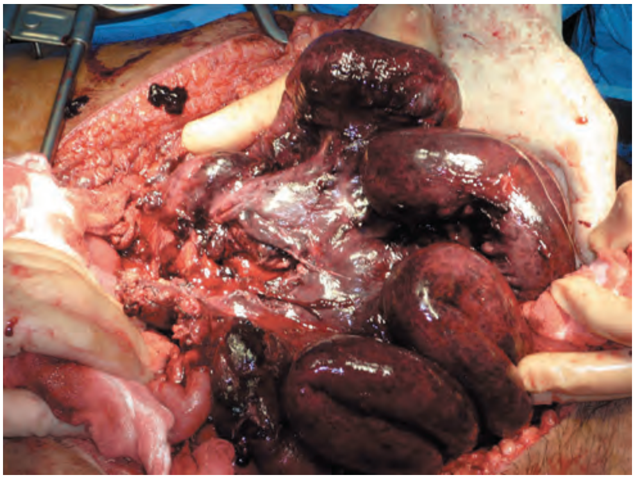


Figure 14.49 Infarction of the small bowel due to venous outflow obstruction after superior mesenteric vein ligation. Note the congested dark appearance of the mesentery, which upon pathological analysis demonstrated sequestered clotted blood within the mesenteric veins.

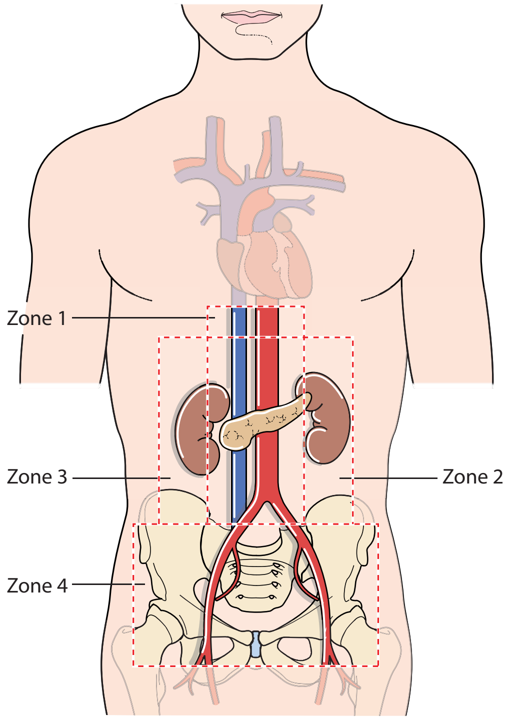


Figure 14.50 Zones of the retroperitoneum. (From Hoyt *et al.*, *Surg Clin North Am* 2001;**81**:1299–330.)

haematomas, both penetrating and blunt, should be explored only if they are expanding or pulsatile, or if they are adjacent to the colon and potentially concealing colonic injury. Zone 3 injuries due to penetrating trauma should be explored to exclude ureteral and iliac artery/vein injury. Those due to blunt trauma usually have associated pelvic fractures and exploration of the haematoma can be hazardous. There is often extensive injury to the rich presacral venous and arterial circulation. Opening the haematoma destroys the tamponade effect, and dissection may result in catastrophic bleeding, and is associated with increased transfusion requirements and high mortality as discrete bleeding points can rarely be identified. If ongoing bleeding is suspected in these cases, pelvic angioembolization should be considered.

Abdominal aorta

Injuries to the abdominal aorta are classified as either suprarenal or infrarenal. Injury to the abdominal aorta is associated with high morbidity and mortality. Shock, acidosis, suprarenal injury and lack of retroperitoneal tamponade are independently associated with mortality. Initial efforts in the management of aortic injuries should be directed at gaining proximal control to limit ongoing haemorrhage. This can be done manually by applying pressure at the diaphragmatic hiatus with either a hand or an aortic compression device. Another option is to clamp the aorta via a left anterolateral thoracotomy; this option should be reserved for patients who are *in extremis* and require open cardiac massage. Once proximal control has been achieved, the suprarenal aorta can be exposed by left-sided medial visceral rotation (Figure 14.51). This should allow visualization of the aorta from its entry at the diaphragmatic hiatus to the bifurcation. Once the injury is identified and isolated, care should be taken to examine both the anterior and posterior walls to evaluate for through-and-through injuries. Small injuries should be repaired primarily. If the defect is large, patch angioplasty can be performed using PTFE or autologous vein. In some instances, significant loss or destruction of the aorta may require replacement of the resected segment with PTFE conduit. If the patient is unstable, temporary intravascular shunts can be used to span the injury and maintain blood flow to the lower limbs. Most evidence suggests systemic heparinization is unnecessary. While temporary intravascular shunts remain a viable and attractive damage control option, indications, type of shunt, dwell time and long-term outcomes have yet to be clarified in the literature.

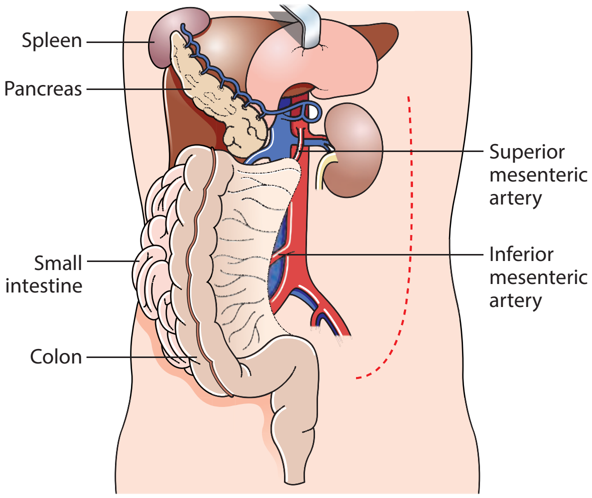


Figure 14.51 A left medial visceral rotation (Mattox manoeuvre) exposing the abdominal aorta from its entry at the diaphragmatic hiatus to the bifurcation including the left iliac vessels. This rotation also provides exposure of the left kidney and renal hilum, as well as the origins of the mesenteric arteries. (From Hoyt *et al.*, *Surg Clin North Am* 2001;**81**:1299–330.)

Inferior vena cava

Inferior vena cava injuries occur after both blunt and penetrating trauma, and are among the most common of intra-abdominal vascular injuries. Inferior vena cava injuries are classified by location as infrarenal, suprarenal or retrohepatic/suprahepatic. Mortality can be high, ranging from 36% to 70%, and increases with increasing number of associated injuries. Mortality is also increased following blunt trauma, if there is release of the retroperitoneal tamponade and in the presence of shock or acidosis.

The most common approach to the infrahepatic inferior vena cava is a right-sided medial visceral rotation (Cattell–Braasch manoeuvre). This manoeuvre exposes the inferior vena cava from the inferior border of the liver to its bifurcation (Figure 14.52). Infrarenal inferior vena cava injuries are considered the easiest to repair, carry the lowest mortality rate and can usually be controlled with compression directly above and below the injury digitally or with sponge sticks. The majority of injuries to the inferior vena cava are repaired primarily. The back wall should be inspected for through-and-through injuries. Posterior defects can be repaired by rotating the vessel, or through the anterior defect. In circumstances in

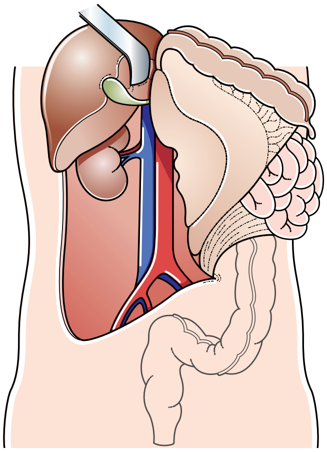


Figure 14.52 Depiction of a right medial visceral rotation (Cattell–Braasch manoeuvre) exposing the right kidney and renal hilum, inferior vena cava, right iliac vessels and the posterior aspect of the duodenum and pancreas. (From Hoyt *et al.*, *Surg Clin North Am* 2001;**81**:1299–330.)

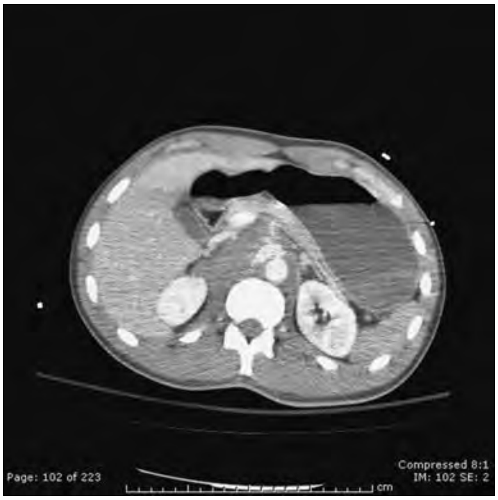


Figure 14.53 Axial CT scan with evidence of coeliac artery pseudoaneurysm. Note the active extravasation of contrast just distal to the origin of the vessel and associated retroperitoneal zone 1 haematoma.

which the defect is large, either saphenous vein or prosthetic patches can be utilized to repair the defect. In damage control situations ligation of the infrarenal inferior vena cava can be considered. In these cases four-compartment lower extremity fasciotomies should be considered as compartment syndrome is common. The bilateral lower extremities should be elevated in all cases in order to decrease accumulation of oedema. In contrast to infrarenal inferior vena cava injuries, the suprarenal inferior vena cava should not be ligated in cases of severe injury as this would preclude renal outflow.

Injuries to the suprarenal or retrohepatic inferior vena cava are particularly difficult to treat, and carry the highest mortality rates. Bleeding is massive and the diagnosis should be suspected when major hepatic parenchymal bleeding persists despite the Pringle manoeuvre. Mobilization and division of the hepatic ligaments are required and vascular control by isolation and clamping of the hepatic inflow and outflow may be necessary. Occlusion of the hepatic veins and inferior vena cava are poorly tolerated owing to the abrupt decrease in venous return to the heart. The use of atriocaval shunts has been described as a temporizing measure to bypass the area of injury while maintaining venous return to the heart until adequate mobilization and repair of the injury is completed. Atriocaval shunts are rarely used, and are associated with mortality exceeding 50%. Another damage control option is total hepatic isolation; requiring clamping of the supra- and infrahepatic inferior vena cava and a Pringle manoeuvre, this also requires application of an aortic cross-clamp to prevent arrest due to severely diminished preload. Total hepatic isolation still often results in arrest or severe shock and outcomes are generally poor. Hepatic venous injury can usually be treated with lateral repair after mobilization and division of the hepatic ligaments but ligation of a single hepatic vein is well tolerated.

Mesenteric vessels

Coeliac artery injuries are rare and primarily due to penetrating trauma (Figure 14.53). Most patients present with hypotension, haemoperitoneum or peritonitis. Mortality associated with coeliac arterial injury is high, and correlates with advancing AAST-OIS grading, hypotension and presence of associated injuries. Exposure of the coeliac artery is best approached via left medial visceral rotation. Once the injury is exposed and proximal and distal control obtained, the surgical options include ligation or repair. Complex reconstructions are not well tolerated and are not recommended. The coeliac artery can be ligated safely if the superior mesenteric artery is patent due to extensive collateral circulation. There are no reports of significant bowel ischaemia following coeliac ligation; however, ischaemia and necrosis of the gallbladder has been described, and cholecystectomy is advocated for patients undergoing ligation. Branches of the coeliac artery may be ligated with relative impunity; however, common hepatic artery ligation distal to the origin of the gastroduodenal artery can result in hepatic ischaemia, particularly if the portal vein is also compromised.

Superior mesenteric artery injuries can be classified according to the schema created by Fullen *et al.* (1972) or by that created by the AAST-OIS. Small defects should be repaired primarily; however, large defects are unlikely to come together primarily as extensive branching of the superior mesenteric artery makes significant mobilization impossible. In these cases, techniques such as vein patch, interposition graft and reimplantation may be utilized. In unstable patients, if primary repair is not possible the injury should be ligated or shunted. Ligation of the superior mesenteric artery is tolerated because of collateral flow through the coeliac axis; however, mortality is higher when ligating proximal than when ligating distal injuries. It should also be kept in mind that bowel ischaemia may result if multiple distal injuries are ligated. In damage control situations temporary intravascular shunts are an alternative to ligation. Compared with extremity injury, truncal temporary intravascular shunts are associated with significantly higher transfusion requirements and higher rates of thrombosis. Strong consideration should be given to temporary abdominal closure for severe superior mesenteric artery injuries because of the high risk of ischaemic complications and abdominal compartment syndrome, particularly after ligation or when concomitant arterial and venous injuries are present. Venous injuries are difficult to control and repair without tension or tearing and are more prone to thrombosis, particularly when prosthetic materials are used.

In very rare cases, patients with superior mesenteric artery injuries may be haemodynamically stable; in these cases, there may be a role for endovascular intervention. A few small case reports of angioembolization for superior mesenteric artery injuries found embolization was successful in controlling haemorrhage in all cases, with no documented bleeding, ischaemia or perforation requiring subsequent laparotomy. However, angiography cannot diagnose or address associated injuries, and it may be best suited to a role as a surgical adjunct.

Injuries to the inferior mesenteric artery/inferior mesenteric vein are rare. Generally, isolated injuries are well tolerated if they are surgically addressed quickly. Mortality is highly dependent on physiological derangements at the time of admission and associated injuries. Inferior mesenteric artery/inferior mesenteric vein injuries are approached directly and, if amenable, primarily repaired or ligated. Ligation is generally well tolerated because of the extensive collateral circulation. However, the surgeon should keep in mind that, although extremely rare, there can be complications of rectal or colonic ischaemia, particularly in patients with advanced atherosclerotic disease.

Renal artery/vein

Renal vascular injuries consist of avulsion, laceration or thrombosis. Penetrating injuries are more common than blunt and most often result in laceration or transection and haemorrhage. Blunt injury can result in the formation of an intimal flap, causing thrombosis and occlusion of the renal artery. Prompt identification of renal artery occlusion is required if an attempt to revascularize the kidney is pursued, as delays to treatment increase the likelihood of nephrectomy. Stable patients should undergo CT scanning to assess for renal or renovascular injury; if unavailable, intravenous pyelography can be performed. Intraoperatively, renovascular injuries may be evidenced by either central or lateral retroperitoneal haematomas. Renal and renal vascular injuries are graded based on the renal AAST-OIS, which has been validated to predict outcome in patients with renal trauma. Advanced grade and blunt mechanisms are both associated with higher nephrectomy rates.

Renal injuries are approached by either right or left medial visceral rotation, the kidney is then delivered from Gerota's fascia. After adequate exposure and proximal control, small renal artery injuries can be repaired primarily. More frequently, renal artery injuries require mobilization and end-to-end repair. Interposition grafts using saphenous vein or PTFE can also be considered for larger arterial defects. Although renal artery repair should be attempted if diagnosed early, renal salvage beyond 6 hours after injury is rare. Nephrectomy should be performed in unstable patients as long as there is confirmation of a contralateral kidney. Autotransplantation or *ex vivo* renal repair may be contemplated in haemodynamically stable patients with bilateral renal injuries to avoid chronic renal failure. Despite an aggressive approach to renal artery repair, it is successful in <30% of cases.

The majority of renal vein injuries can be repaired primarily. Injury to the left renal vein can be treated safely with ligation near the inferior vena cava provided that adrenal and gonadal collaterals have been preserved. The right renal vein lacks collateral outflow, and nephrectomy is indicated if repair of the vein is not possible.

Iliac artery and vein

Injury to these vessels is most commonly due to penetrating trauma, and can be quite morbid. Injuries due to blunt trauma, although rarer, are associated with increased morbidity and mortality, usually because of the high number of associated injuries seen in this population. Patients with shock and evidence of significant haemoperitoneum should be taken for operative exploration. Injury to the iliac artery should be suspected with rapidly expanding or pulsatile haematomas in zone 3, or in any patient with penetrating pelvic trauma and a zone 3 haematoma. Exposure of the iliac artery is achieved by incising the retroperitoneum over the aortic bifurcation, alternatively left or right medial visceral rotation may be performed to expose the left or right iliacs. Initial proximal control is achieved with manual compression; care should be taken to identify and preserve the ureter during dissection. Vessel loops or vascular clamps can then be used to gain definitive proximal control. Distal control of the external iliac can be obtained near the inguinal ligament. Control of the internal iliac artery is also necessary to avoid back bleeding from the extensive collateral circulation during isolation and repair. Small injuries should be repaired primarily; larger or destructive injuries should be resected. The resected gap can be spanned with primary anastomosis after adequate mobilization of the proximal and distal ends or with PTFE or saphenous vein interposition grafts. Injuries to the iliac arteries may be associated with injury to the bowel resulting in the spillage of enteric contents into the pelvis; in these cases, ligation of the injury with subsequent extra-anatomic bypass in a non-contaminated field may be required. In patients with significant physiological derangement and complex injuries, shunting is an alternative to ligation. In common and external iliac arterial injury there are some data to suggest that temporary intravascular shunts can significantly reduce the need for later fasciotomies and amputation. Injuries to the internal iliac artery can be ligated with little consequence owing to the rich pelvic collateral blood flow.

Among haemodynamically stable patients, diagnosis of iliac arterial injury primarily occurs following CT scan, which has a sensitivity of >90% for detecting pelvic bleeding requiring intervention. If associated hollow viscous injury cannot be ruled out, or if the patient's clinical course deteriorates, immediate laparotomy and surgical repair is the treatment of choice. However, in stable patients with isolated iliac artery injuries, good outcomes have been noted with the deployment of covered stent grafts. Successful occlusion of the injury and haemostasis are achieved in 75–100% of cases, with few complications. The most common complications are stent–graft occlusion followed by stenosis.

Iliac vein injuries are frequently associated with arterial injuries. Iliac vein injury can occur after blunt or penetrating trauma and as a result of iatrogenic injury following pelvic procedures. Mortality for venous injuries is similar to that for arterial injuries and can be as high as 70%. Access to common iliac vein injuries, especially on the right side, may be limited by the overlying common iliac artery. Transection of the artery for visualization and repair of the venous injury has been described historically, but is not advocated in modern times. Minor lacerations can be repaired primarily; however, more destructive injuries associated with gunshot wounds and blunt trauma most often require ligation. Complex venous reconstructions after traumatic injury are rarely indicated and should only be performed in stable patients. Postoperative anticoagulation should be considered if the venous repair results in significant narrowing of the vein to prevent thrombosis. Previous studies have shown that ligation of the common or external iliac vein is well tolerated with few adverse sequelae. Complications may include extremity oedema, compartment syndrome, thromboembolic complications and outflow ischaemia. While leg oedema is common after ligation, compartment syndrome is rare unless there is also arterial injury or prolonged hypotension.

Superior mesenteric vein/portal vein

Injuries to the portal and superior mesenteric veins are rare, highly lethal injuries occurring in <1% of all traumas. Most injuries occur as a result of penetrating trauma, lead to exsanguination in a short period of time, and present with hypotension or peritonitis requiring emergency laparotomy. An additional minority may require resuscitative thoracotomy. The need to undergo either resuscitative or operating theatre thoracotomy is associated with an increased risk of death.

Because the portal vein and superior mesenteric vein are centrally located, nearly all patients have associated injuries, averaging three or more per patient, and the surgical approach can be difficult owing to the proximity of other vital structures. Initial haemorrhage control of injuries in the hepatoduodenal ligament can be obtained by application of the Pringle manoeuvre. Once the haematoma and areolar tissue in the hepatoduodenal ligament have been dissected, vascular clamps or vessel loops should be placed above and below the injury. The distal portal vein is exposed by medial retraction of the common bile duct; however, approach to the more proximal portal vein requires reflection of the pancreas and duodenum via a wide Kocher or Cattell–Braasch manoeuvre. Extra length and access to more proximal portal and distal superior mesenteric vein injuries can be gained by dividing the neck of the pancreas. Portal vein injuries can be addressed by repair, reanastomosis, interposition graft, portosystemic shunt or ligation. Ligation is well tolerated as long as the hepatic artery is patent, with mortality ranging from 20% to 90%; however, concomitant hepatic artery injury mandates portal vein repair. Ligation of the portal vein may result in splanchnic hypervolaemia, and systemic hypovolaemia causing haemodynamic compromise. Careful monitoring of preload and aggressive intravascular volume replacement are essential in these circumstances.

Superior mesenteric vein injuries can be treated with ligation or repair with similar outcomes. If ligation is performed, it should be kept in mind that, similar to portal vein ligation, the syndrome of splanchnic hypervolaemia/systemic hypovolaemia is a significant risk. Because of the risk of bowel ischaemia due to acute outflow occlusion and portal hypertension, damage control laparotomy is particularly well suited to portal vein and superior mesenteric vein ligation. A second look to assess bowel viability is useful and the risks of abdominal compartment syndrome may be decreased with temporary abdominal closure techniques.

Injuries to the extremities

Extremity vascular injuries are common sequelae of both blunt and penetrating trauma. Lower extremity injuries are more common than upper, and the majority of injuries are due to penetrating trauma. Although tolerance to ischaemia is greater than that of the viscera or brain, rapid diagnosis and treatment are still essential. Because of the superficial location of extremity vessels clinical examination is very sensitive for detection of arterial injury. Patients with any hard signs of injury should be taken to surgery immediately. The one exception would be in blunt trauma patients or victims of multilevel penetrating trauma, such as shotgun wounds, where angiography or CT-A can be used for surgical planning to optimize exposure and exploration.

In stable patients without hard signs of injury a complete assessment of all peripheral pulses should be performed. If any pulse asymmetry or obvious orthopaedic injury (Box 14.2) is noted, an ABI or BBI should be performed. Patients with no hard or soft signs of injury, and normal ABI/BBI results may be discharged. Patients with ABI/BBI <0.9 should undergo imaging. Options include ultrasound, angiography and CTA.

Ultrasound is an excellent option below the inguinal ligament and below the shoulder. It has the benefit of being portable, non-invasive and does not require ionizing radiation. However, fractures, subcutaneous emphysema and soft-tissue injury significantly decrease sensitivity. Traditional angiography, including surgeon-performed on-table angiography, is both sensitive and specific with a low complication rate. Because both ultrasound and angiography require mobilization of advanced imaging teams with specialized training and are unlikely to be immediately available in most centres after hours or on holidays/weekends, CT-A has become popular as a screening method. It is rapid, easily performed, does not require an interventional radiology team or arterial puncture and has few complications. CT-A also has the benefit of giving additional information on

BOX 14.2 Orthopaedic injuries associated with vascular trauma

- Posterior knee dislocation
- Distal femur
- Proximal tibia
- Supracondylar fracture of the humerus
- Clavicular fracture
- Shoulder dislocation
- First rib fracture

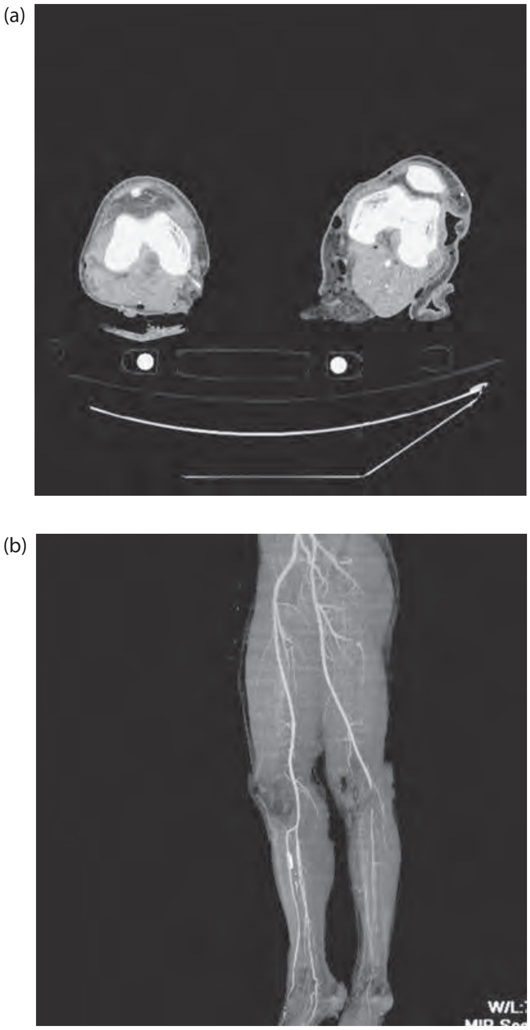


Figure 14.54 CT scan demonstrating severe bilateral lower extremity blunt trauma. (a) Axial views demonstrate significant soft-tissue injury and loss of the left knee, and occlusion of the right popliteal artery. (b) Three-dimensional reconstruction of the same patient demonstrating a long segment occlusion of the right popliteal artery with distal reconstitution of flow.

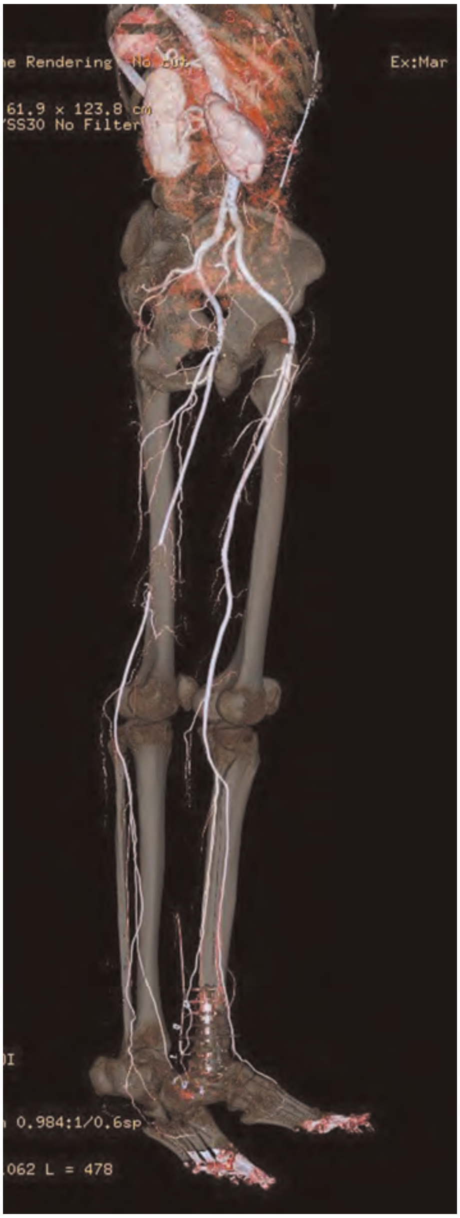


Figure 14.55 Three-dimensional CT angiography reconstruction of a patient with injury and occlusion to the left distal superficial femoral artery just above Hunter's canal.

non-vascular structures such as the bone, joint and soft tissue (Figure 14.54a,b). Overall CT-A has excellent sensitivity and specificity for detecting both upper and lower extremity vascular injuries. Accuracy appears to be particularly good when using multidetector 16- and 64-slice scanners, and reformatting such as maximum intensity projections and two- and three-dimensional reconstruction (Figure 14.55). CT-A is also reported as being the least uncomfortable vascular examination in many series comparing it with MRA, ultrasound and traditional angiography, and may have a cost benefit over traditional angiography.

Limitations of CT-A include the need to transport the patient to the radiology suite and exposure to contrast and ionizing radiation. Additionally, in victims of gunshot and shotgun wounds images may be compromised by metallic fragments.

The femoral artery is the most commonly injured vessel in the lower extremity. Although most often due to penetrating trauma, fractures, dislocations and iatrogenic injuries are also important causes. Femoral artery injuries are often associated with injury to accompanying veins and nerves. In addition to arterial repair, extensive attempts at deep vein reconstruction are warranted to minimize early and late morbidity. In general, injuries to the profunda femoris artery should also be repaired because of the significant incidence of subsequent atherosclerotic vascular disease involving the superficial femoral artery. Injuries are repaired by resection and primary anastomosis or interposition graft. The contralateral saphenous vein is preferentially used when a venous conduit is required for major lower extremity arterial and venous injuries.

The popliteal vessels are particularly susceptible to blunt traumatic injury from fractures and dislocations. The amputation rate for unreconstructed injuries in this area is as high as 61%. Concomitant popliteal vein injury decreases the likelihood of success following arterial reconstruction. Therefore, expeditious repair of both the vein and artery is essential.

Although the ideal sequence of repair in patients with combined orthopaedic and popliteal vascular injuries has been debated for many years, most surgeons now use shunts to temporarily perfuse the extremity while orthopaedic stabilization is performed (Figure 14.56). Proximal and distal thromboembolectomy should be performed prior to shunt placement and local heparinization should be used if possible. Temporary intravascular shunts appear to be well tolerated, particularly in the extremities, and do not require systemic heparinization. Maximum dwell times are unknown, but some series have shown excellent patency as long as 72 hours. Once the fractures have been stabilized the temporary intravascular shunts can be removed and definitive vascular repair performed in the stabilized limb.

Injury to the small vessels below the popliteal trifurcation can pose difficult management problems. When both tibial arteries are injured, the amputation rate may reach 65%, whereas with injury to only one vessel the rate is much lower. Arteriography in this area is important in defining the site and extent of all the injuries. This is particularly true with multiple pellet wounds from a shotgun blast. If arteriography demonstrates only a single isolated arterial injury below the knee without active bleeding, pseudoaneurysm or arteriovenous fistula, and the patient does not have an ischaemic limb, then observation without surgical intervention is justified. On the other hand, any patient who demonstrates ischaemia, bleeding or a vascular complication of the injury should undergo immediate exploration and repair of the artery. If a vein graft is required for repair, it should be harvested from the contralateral extremity. Fasciotomy is often required when prolonged ischaemia or soft-tissue injury is present. Salvage of a useful limb is often more dependent on the extent of neuromuscular injury than vascular injury.

Vascular injuries of the upper extremity are extremely common. These injuries are not only seen with penetrating trauma but are also encountered after use of the brachial artery for arteriographic and cardiac catheterization, and following cannulation for arterial blood gas monitoring. Long-bone fractures with or without joint dislocation are another potential source of vascular trauma. In children the brachial artery is particularly vulnerable to injury from supracondylar fractures of the humerus. Most upper extremity vascular injuries do not result in limb-threatening ischaemia and physical examination or angiography can establish the diagnosis. Injuries to the axillary and brachial arteries require urgent exploration and can usually be managed by thrombectomy and primary repair with or without a vein patch. Injuries to the radial and ulnar arteries are often easily repaired but either can be ligated if the other vessel is patent and there is an intact palmar arch. Fasciotomy of the forearm may be needed if ischaemia or soft-tissue injury causes a compartment syndrome.

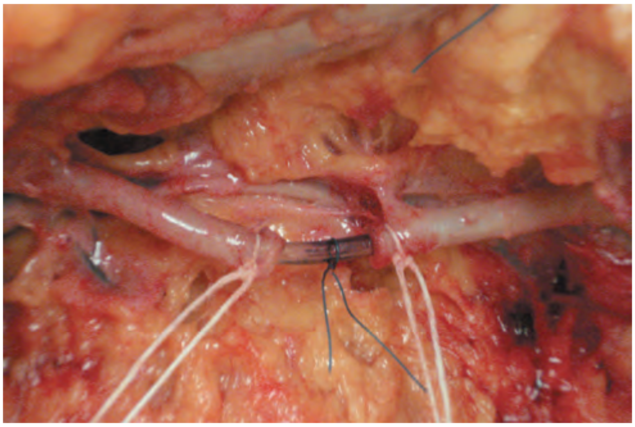


Figure 14.56 A temporary intravascular shunt spanning a defect in the popliteal artery; note that a silk suture has been used to mark the midline of the shunt to prevent dislodgement during orthopaedic manipulation of the knee joint.

Acute and chronic sequelae of vascular injury

Haemorrhage

Bleeding subsequent to arterial repair is uncommon, occurring in <5% of cases. It may present as frank haemorrhage from the operative incision or as a rapidly expanding haematoma. In virtually all cases, the cause is a technical error, resulting from inadequate haemostasis in the wound or from the suture line of the vascular repair. Rarely, bleeding may be due to a missed arterial or venous injury. Patients who develop this complication should be returned to the operating theatre immediately for evacuation of any surrounding haematoma and identification and correction of the problem.

Thrombosis

Occlusion of a vascular repair is the most frequently encountered acute complication of vascular surgery. Depending on the vessel involved and the type of repair (i.e. lateral arteriorrhaphy, interposition grafting, etc.) thrombosis can occur in approximately 10% of cases. If late thromboses are included, the incidence rises to about 20%. Early thrombosis is due to a technical problem, most commonly stenosis of the repair or inadequate thrombectomy. Other causes include intimal dissection with prolapse and missed injury. Because these technical problems are easily corrected if discovered early, it is prudent to perform completion arteriography or intraoperative ultrasound if available.

Thrombosis presents with a loss or diminution of distal pulses; these may or may not be associated with signs of ischaemia (i.e. pallor, poikilothermia, pain, paralysis and paraesthesia). If loss of pulse occurs the diagnosis is thrombosis and there is no need for confirmatory arteriography. These patients should be returned as soon as possible to the operating theatre for thrombectomy, possible revision of the repair and completion arteriography. On the other hand, if pulses were never palpable after the repair and the patient is hypothermic, vasospasm may be present. In such cases, every effort should be made to rewarm the patient and improve perfusion to the extremity. If pulses do not return, either arteriography or exploration should be considered.

Arteriovenous fistula

Arteriovenous fistula after trauma occurs in 2–7% of cases. The most common causes are fragment, shrapnel or shotgun wounds. Numerous physiological changes occur as flow increases through the fistula – going from the high-pressure arterial system to the low-resistance venous system. Among these changes are increases in cardiac output, heart rate, central venous pressure and blood volume. Local changes in the region of the fistula include dilatation of the proximal artery and the distal vein, distal varicosities and extremity oedema. Most of these changes are reversible with correction of the fistula. Patients may present with signs of venous stasis disease, arterial insufficiency or heart failure. The physical findings are often suggestive of the diagnosis and include a palpable thrill and audible bruit or a continuous machinery-type murmur. Occlusion of the fistula may cause a decrease in the elevated pulse rate (Nicoladoni–Branham sign). Arteriography should always be performed to confirm the diagnosis, localize the lesion anatomically and determine the presence of any associated pathology. Less than 2% of fistulas will close spontaneously, so treatment is required. Treatment options include endovascular placement of covered stent grafts and open repair.

Pseudoaneurysm

The incidence of pseudoaneurysm following trauma is difficult to ascertain since it is often reported in association with arteriovenous fistulas. A pseudoaneurysm may cause symptoms by encroachment on local structures as it expands or it may present acutely with rupture, distal embolization or thrombosis. Small pseudoaneurysms (<3 cm) will often thrombose spontaneously, particularly when due to iatrogenic injury. In stable patients without local complications, a period of observation is reasonable. This may be augmented with percutaneous treatments such as ultrasound-guided compression, image-guided thrombin injection, or endovascular techniques such as coil or gel embolization in aneurysms with long narrow necks or occlusion of the mouth of the pseudoaneurysm with a covered stent graft.

If the pseudoaneurysm is large or persistent, arteriography should be obtained to define the anatomy and associated pathology, and followed by surgical repair. Although primary repair may be possible, most cases require resection with end-to-end anastomosis or interposition graft.

Infection

Infection of an arterial suture line is a disastrous complication, often leading to haemorrhage, thrombosis, distal ischaemia and eventual amputation. Factors predisposing to infection include closure of a contaminated wound, inadequate soft-tissue coverage of an arterial repair and inadequate debridement of a traumatized, contaminated vessel. This complication is best avoided by vigorous cleansing of contaminated wounds, use of appropriate antibiotics, aggressive debridement of devitalized tissues and coverage of arterial repairs with well-vascularized soft tissue.

Compartment syndrome

Extremity compartment syndrome is the result of soft-tissue trauma or ischaemia–reperfusion injury, which leads to swelling within a closed fascial space. This contained swelling results in a rise in tissue hydrostatic pressure to the point that blood flow is compromised. If untreated, compartment syndrome will result in myonecrosis and limb dysfunction. The most commonly involved areas are the anterior compartment in the lower leg and the volar compartment in the forearm. Since nerve tissue is more susceptible to ischaemia than is muscle, initial symptoms include paraesthesia and pain in the involved extremity. The muscle is usually tense and pain may be severe, especially with passive movement of the involved compartment. Pulses are usually palpable, even in advanced stages, and are not a reliable indicator of the severity of the syndrome. If one suspects development of a compartment syndrome, compartment pressure should be measured. This can be done by insertion of a plastic cannula or a wick catheter directly into the involved muscle or muscle group. Fasciotomy is indicated in those patients with a compartmental pressure >40 mmHg initially or a pressure of 30 mmHg that is sustained for longer than 4 hours. If compartmental pressures cannot be obtained, one must proceed based upon the clinical impression. Risk factors associated with compartment syndrome include crush injuries, concomitant arterial and venous injury, vascular injury requiring ligation, and ischaemic time ≥6 hours.

Treatment of compartment syndrome should not be delayed, as irreversible myonecrosis with resultant contracture will occur within 12 hours. Treatment of compartment syndrome of the calf consists of four-compartment fasciotomy; this can be performed through a standard two-incision approach or through a single lateral incision (Figure 14.57). Compartment syndrome of the thigh requires a single lateral incision to decompress the three compartments of the thigh. The upper and lower arm contain two and three compartments, respectively, and can be decompressed via a single anterolateral oblique incision running the length of the upper extremity, supplemented with an additional posterior incision in the forearm to access the posterior compartment and mobile wad. Care should be taken to open the carpal tunnel during forearm fasciotomy.

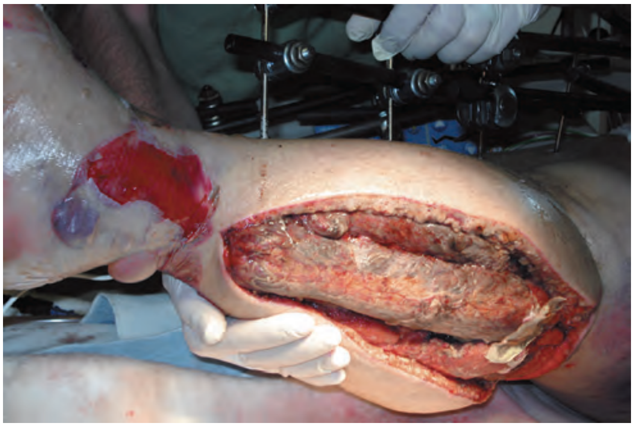


Figure 14.57 A four-compartment calf fasciotomy. The image depicts the lateral incision; note the bulging muscle beds. An external fixator spanning an associated severe tibia/fibula fracture can be seen in the background.

GUIDE TO FURTHER READING

General considerations

Bansal V, Fortlage D, Lee JG, *et al.* Hemorrhage is more prevalent than brain injury in early trauma deaths: the golden six hours. *Eur J Trauma Emerg Surg* 2009;**35**:26–30.

Trunkey DD. Trauma. Accidental and intentional injuries account for more years of life lost in the U.S. than cancer and heart disease. Among the prescribed remedies are improved preventive efforts, speedier surgery and further research. *Sci Am* 1983;**249**:28–35.

Metabolic response to trauma

Junger WG, Hoyt DB, Lieu FC, *et al.* Immunosuppression after endotoxin shock: the result of multiple anti-inflammatory factors. *J Trauma* 1996;**40**:702–709.

Latifi R. Nutritional therapy in critically ill and injured patients. *Surg Clin North Am* 2011;**91**:579–593.

Naplitano LM, Faist E, Wichmann MW, Coimbra R. Immune dysfunction in trauma. *Surg Clin North Am* 1999;**79**:1385–1416.

Smith JW, Gamelli RL, Jones SB, Shankar R. Immunologic response to critical injury and sepsis. *J Intensive Care Med* 2006;**21**:160–172.

Prehospital care and transport

American College of Surgeons Committee on Trauma. *Resources for the Optimal Care of the Injured Patient*. Chicago, IL: ACS, 2006.

Cox S, Smith K, Currell A, *et al.* Differentiation of confirmed major trauma patients and potential major trauma patients using prehospital trauma triage criteria. *Injury* 2011;**42**:889–895.

Purtill MA, Benedict K, Hernandez-Boussard T, *et al.* Validation of a prehospital trauma triage tool: a 10-year perspective. *J Trauma* 2008;**65**:1253–1257.

Assessment of the injured patient

Ball CG, Williams BH, Wyrzykowski AD, *et al.* A caveat to the performance of pericardial ultrasound in patients with penetrating cardiac wounds. *J Trauma* 2009;**67**:1123–1124.

Bickell WH, Wall MJ Jr, Pepe PE, *et al.* Immediate versus delayed fluid resuscitation for hypotensive patients with penetrating torso injuries. *N Engl J Med* 1994;**331**:1105–1109.

Bratton SL, Chestnut RM, Ghajar J, *et al*; Brain Trauma Foundation, American Association of Neurological Surgeons, Congress of Neurological Surgeons, Joint Section on Neurotrauma and Critical Care, AANS/CNS. Guidelines for the management of severe traumatic brain injury. I. Blood pressure and oxygenation. *J Neurotrauma* 2007;**24**(Suppl. 1):S7–13.

Demetriades D, Sava J, Alo K, *et al.* Old age as a criterion for trauma team activation. *J Trauma* 2001;**51**:754–6; discussion 756–757.

EAST Practice Parameter Workgroup for Pre-hospital Fluid Resuscitation, Eastern Association for the Surgery of Trauma. *Practice Management Guidelines for Prehospital Fluid Resuscitation in the Injured Patient*. Available from: http://www.east.org/tpg/FluidResus.pdf.

Fraga GP, Bansal V, Coimbra R. Transfusion of blood products in trauma: an update. *J Emerg Med* 2010;**39**:253–260.

Holcomb JB, Wade CE, Michalek JE, *et al.* Increased plasma and platelet to red blood cell ratios improves outcome in 466 massively transfused civilian trauma patients. *Ann Surg* 2008;**248**:447–458.

Junger WG, Coimbra R, Liu FC, *et al.* Hypertonic saline resuscitation: a tool to modulate immune function in trauma patients? *Shock* 1997;**8**:235–241.

Rhee PM, Acosta J, Bridgeman A, *et al.* Survival after emergency department thoracotomy: review of published date from the past 25 years. *J Am Coll Surg* 2000;**190**:288–298.

Riskin DJ, Tsai TC, Riskin L, *et al.* Massive transfusion protocols: the role of aggressive resuscitation versus product ratio in mortality reduction. *J Am Coll Surg* 2009;**209**:198–205.

Rozycki GS, Feliciano DV, Ochsner MG, *et al.* The role of ultrasound in patients with possible penetrating cardiac wounds: a prospective multicenter study. *J Trauma* 1999;**46**:543–551; discussion 551–2.

Scalea TM, Simon HM, Duncan AO, *et al.* Geriatric blunt multiple trauma: improved survival with early invasive monitoring. *J Trauma* 1990;**30**:129–34; discussion 134–136.

Working Group, Ad Hoc Subcommittee on Outcomes, American College of Surgeons–Committee on Trauma. Practice management guidelines for emergency department thoracotomy. *J Am Coll Surg* 2001;**193**:303–309.

Neck

Berne JD, Cook A, Rowe SA, *et al.* A multivariate logistic regression analysis of risk factors for blunt cerebrovascular injury. *J Vasc Surg* 2010;**51**:57–64.

Biffl WJ, Cothren C, Moore EE, *et al.* Western Trauma Association critical decision in trauma: screening for and treatment of blunt cerebrovascular injuries. *J Trauma* 2009;**67**:1150–1153.

Britt LD. Penetrating neck trauma. In: Fischer JE, Bland KI (eds) *Mastery of Surgery*, 5th edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2007:381–386.

Britt LD, Weireter LJ, Cole FJ. Management of acute neck injuries. In: Feliciano DV, Mattox KL, Moore EE (eds) *Trauma*, 6th edn. New York, NY: McGraw Hill, 2008:467–477.

Bromberg WJ, Collier BC, Diebel LN, *et al.* Blunt cerebrovascular injury practice management guidelines: the Eastern Association for the Surgery of Trauma. *J Trauma* 2010;**68**:471–477.

Case SJ, de Alwis WD. Emergency department assessment and management of stab wounds to the neck. *Emerg Med Aust* 2010;**22**:201–210.

Gale SC, Mattox KL. Trauma to the head, face, and neck. In: Gracias VH, Reilly PM, McKenney MG, Velmahos GC (eds) *Acute Care Surgery: A Guide for General Surgeons*. New York, NY: McGraw Hill, 2009:99–108.

Koletsis E, Prokakis C, Baltayiannis N, *et al.* Surgical decision making in tracheobronchial injuries on the basis of clinical evidences and the injury's anatomical setting: a retrospective analysis. *Injury* 2012;**43**: 1437–1441.

Osborn T, Bell RB, Qaisi W, *et al.* Computed tomographic angiography as an aid to clinical decision making in selective management of penetrating injuries to the neck: a reduction in the need for operative exploration. *J Trauma* 2008;**64**:1466–71.

Wu, JT, Mattox KL, Wall MJ. Esophageal perforations: new perspectives and treatment paradigms. *J Trauma* 2007;**63**:1173–84.

Thorax

Bulger EM, Klotz P, Jurkovich GJ. Epidural anesthesia improves outcomes after multiple rib fractures. *Surgery* 2004;**136**:426–30.

Cohn SM, DuBose JJ. Pulmonary contusions: an update on recent advances in clinical management. *World J Surg* 2010;**34**:1959–70.

Holcomb JB, McMulin NR, Kozar RA, *et al.* Morbidity from rib fractures increases after age 45. *J Am Coll Surg* 2003;**196**:549–55.

Hunt I, Thakar C, Southon R, *et al.* Establishing a role for intra-pleural fibrinolysis in managing traumatic hemothoraces. *Interact J Cardiovasc Thoracic Surg* 2009;**8**:129–33.

Kimbrell BJ, Yamzon J, Petrone P, *et al.* Intrapleural thrombolysis for management of undrained traumatic hemothorax: a prospective observational study. *J Trauma* 2007;**62**:1175–9.

Livingston DH, Hauser CJ. Chest wall and lung. In: Feliciano DV, Mattox KL, Moore EE (eds) *Trauma*, 6th edn. New York, NY: McGraw Hill, 2008:525–52.

Sirmali M, Turut H, Topcu S, *et al.* A comprehensive analysis of traumatic rib fractures: morbidity, mortality and management. *Eur J Cardiothorac Surg* 2003;**24**:133–8.

Abdomen

Bansal V, Conroy C, Tominaga GT, *et al.* The utility of seat belt signs to predict intra-abdominal injury following motor vehicle crashes. *Traffic Injury Prevention* 2009;**10**:567–72.

Barker DE, Kauffman HJ, Smith LA, *et al.* Vacuum pack technique of temporary abdominal closure: a 7-year experience with 112 patients. *J Trauma* 2000;**48**:201–6.

Burch JM, Feliciano DV, Mattox KL. Colostomy and drainage for civilian rectal injuries: is that all? *Ann Surg* 1989;**209**:600–10.

Cheatham MK, Malbrain MLNG, Kirkpatrick A, *et al.* Results from the international conference of experts on intra-abdominal hypertension and abdominal compartment syndrome. II. Recommendations. *Intensive Care Med* 2007;**33**:951–62.

Coburn, M. Genitourinary trauma. In: Feliciano DV, Mattox KL, Moore EE (eds) *Trauma*, 6th edn. New York, NY: McGraw Hill, 2008:789–825.

Cothren CC, Osborn PM, Moore EE, *et al.* Preperitoneal pelvic packing for hemodynamically unstable pelvic fractures: a paradigm shift. *J Trauma* 2007;**62**:834–9; discussion 839–42.

Demetriades D, Murray JA, Chan L, *et al.* Penetrating colon injuries requiring resection: diversion or primary anastomosis? An AAST prospective multicenter study. *J Trauma* 2001;**50**:765–75.

Gonzalez RP, Falimirski ME, Holevar MR. The role of presacral drainage in the management of penetrating rectal injuries. *J Trauma* 1998;**45**:656–61.

Gonzalez RP, Phelan H, Hassan M, *et al.* Is fecal diversion necessary for nondestructive penetrating extraperitoneal rectal injuries? *J Trauma* 2006;**61**:815–19.

Hadjizacharia P, Inaba K, Teixeira PG, *et al.* Evaluation of immediate endoscopic realignment as a treatment modality for traumatic urethral injuries. *J Trauma* 2008;**64**:1449–50.

Jacoby RC, Wisner DH. Injury to the spleen. In: Feliciano DV, Mattox KL, Moore EE (eds) *Trauma*, 6th edn. New York, NY: McGraw Hill, 2008:661–80.

Murray JA, Demetriades D, Asensio JA, *et al.* Occult injuries to the diaphragm: prospective evaluation of laparoscopy in penetrating injuries to the lower left chest. *J Am Coll Surg* 1998;**187**:626–30.

Patton JH, Lyden SP, Croce MA, *et al.* Pancreatic trauma: a simplified management guideline. *J Trauma* 1997;**43**:234–9.

Rotondo MF, Schwab W, McGonigal MD, *et al.* Damage control: an approach for improved survival in exsanguinating penetrating abdominal injury. *J Trauma* 1993;**35**:375–82.

Schrock T, Blaisdell FW, Mathewson C Jr. Management of blunt trauma to the liver and hepatic veins. *Arch Surg* 1968;**96**:698–704.

Shatz DV, Schinsky MF, Pais LR, *et al.* Immune responses of splenectomized trauma patients to the 23-valent pneumococcal polysaccharide vaccine at 1 versus 7 versus 14 days after splenectomy. *J Trauma* 1998;**44**:760–5.

Sokolove PE, Kuppermann N, Holmes JF. Association between the 'seat belt sign' and intra-abdominal injury in children with blunt torso trauma. *Acad Emerg Med* 2005;**12**:808–13.

Starnes M, Demetriades D, Hadjizacharia P, *et al.* Complications following renal trauma. *Arch Surg* 2010;**145**:377–81.

Stein DJ, Scalea TM. Nonoperative management of spleen and liver injuries. *J Intensive Care Med* 2006;**21**:296–304.

Vasquez JC, Coimbra R, Hoyt DB, Fortlage D. Management of penetrating pancreatic trauma: an 11 year experience of a level-1 trauma center. *Injury* 2001;**32**:753–9.

Velmahos GC, Chahwan S, Hanks SE, *et al.* Angiographic embolization of bilateral internal iliac arteries to control life-threatening hemorrhage after blunt trauma to the pelvis. *Am Surg* 2000;**66**:858–62.

Young JWR, Resnik CS. Fracture of the pelvis: current concepts of classification. *AJR Am J Roentgenol* 1990;**155**:1169–75.

Vascular

Asensio JA, Chahwan S, Hanpeter D, *et al.* Operative management and outcome of 302 abdominal vascular injuries. *Am J Surg* 2000;**180**:528–33; discussion 533–44.

Asensio JA, Petrone P, Roldan G, *et al.* Analysis of 185 iliac vessel injuries: risk factors and predictors of outcome. *Arch Surg* 2003;**138**:1187–93; discussion 1193–4.

Ball CG, Feliciano DV. Damage control techniques for common and external iliac artery injuries: have temporary intravascular shunts replaced the need for ligation? *J Trauma* 2010;**68**:1117–20.

Coimbra R, Prado PA, Araujo LH, *et al*. Factors related to mortality in inferior vena cava injuries. A 5 year experience. *Int Surg* 1994;**79**:138–41.

Demetriades D, Theodorou D, Cornwell E, *et al*. Evaluation of penetrating injuries of the neck: prospective study of 223 patients. *World J Surg* 1997;**21**:41–7; discussion 47–8.

Demetriades D, Velmahos GC, Scalea TM, *et al*.; American Association for the Surgery of Trauma Thoracic Aortic Injury Study Group. Operative repair or endovascular stent graft in blunt traumatic thoracic aortic injuries: results of an American Association for the Surgery of Trauma Multicenter Study. *J Trauma* 2008;**6**4:561–70; discussion 570–1.

Dennis JW, Frykberg ER, Veldenz HC, *et al*. Validation of nonoperative management of occult vascular injuries and accuracy of physical examination alone in penetrating extremity trauma: 5 – to 10-year follow-up. *J Trauma* 1998;**44**:243–52; discussion 242–3.

Dente C, Feliciano D. Abdominal vascular injury. In: Feliciano DV, Mattox KL, Moore EE (eds) *Trauma*, 6th edn. New York, NY: McGraw Hill, 2008:737–757.

DuBose J, Inaba K, Barmparas G, *et al*. Bilateral internal iliac artery ligation as a damage control approach in massive retroperitoneal bleeding after pelvic fracture. *J Trauma* 2010;**69**:1507–14.

du Toit DF, Lambrechts AV, Stark H, Warren BL. Long-term results of stent graft treatment of subclavian artery injuries: management of choice for stable patients? *J Vasc Surg* 2008;**47**:739–43.

Dyer D, Moore EE, Ilke DN, *et al*. Thoracic aortic injury: how predictive is mechanism and is chest computed tomography a reliable screening tool? A prospective study of 1,561 patients. *J Trauma* 2000;**48**:673–82; discussion 682–3.

Fabian TC, Davis KA, Gavant ML, *et al*. Prospective study of blunt aortic injury: helical CT is diagnostic and antihypertensive therapy reduces rupture. *Ann Surg* 1998;**227**:666–76; discussion 676–7.

Fraga GP, Bansal V, Fortlage D, *et al*. A 20-year experience with portal and superior mesenteric venous injuries: has anything changed? *Eur J Vasc Endovasc Surg* 2009;**37**:87–91.

Fullen WD, Hunt J, Altemeier WA. The clinical spectrum of penetrating injury to the superior mesenteric arterial circulation. *J Trauma* 1972;**12**:656–64.

Hoyt DB, Coimbra R, Potenza BM, *et al*. Anatomic exposures for vascular injuries. *Surg Clin North Am* 2001;**81**:1299–330.

Lee WA, Matsumura JS, Mitchell RS, *et al*. Endovascular repair of traumatic thoracic aortic injury: clinical practice guidelines of the Society for Vascular Surgery. *J Vasc Surg* 2001;**53**:187–92.

Levy BA, Zlowodzki MP, Graves M, Cole PA. Screening for extremity arterial injury with the arterial pressure index. *Am J Emerg Med* 2005;**23**:689–95.

Mattox KL, Feliciano DV, Burch J, *et al*. Five thousand seven hundred sixty cardiovascular injuries in 4459 patients. Epidemiologic evolution 1958 to 1987. *Ann Surg* 1989;**209**:698–705; discussion 706–7.

Pappas PJ, Haser PB, Teehan EP, *et al*. Outcome of complex venous reconstruction in patients with trauma. *J Vasc Surg* 1997;**25**:398–404.

Pinto A, Niola R, Tortora G, *et al*. Role of multidetector-row CT in assessing the source of arterial haemorrhage in patients with pelvic vascular trauma. Comparison with angiography. *Radiol Med* 2010;**115**:648–67.

Reilly PM, Rotondo MF, Carpenter JP, *et al*. Temporary vascular continuity during damage control: intraluminal shunting for proximal superior mesenteric artery injury. *J Trauma* 1995;**39**:757–60.

Saletta S, Lederman E, Fein S, *et al*. Transesophageal echocardiography for the initial evaluation of the widened mediastinum in trauma patients. *J Trauma* 1995;**39**:137–41; discussion 141–2.

Velmahos GC, Toutouzas KG, Vassiliu P, *et al*. A prospective study on the safety and efficacy of angiographic embolization for pelvic and visceral injuries. *J Trauma* 2002;**53**:303–8; discussion 308.

White R, Krajcer Z, Johnson M, *et al*. Results of a multicenter trial for the treatment of traumatic vascular injury with a covered stents. *J Trauma* 2006;**60**:1189–95; discussion 1195–6.

Willmann JK, Wildermuth S. Multidetector-row CT angiography of upper- and lower-extremity peripheral arteries. *Eur Radiol* 2005;**15**(Suppl. 4):D3–9.

CHAPTER 15

Head injuries: Pathophysiology and management

SAM ELJAMEL

Introduction

The goals of head injury management are prevention of secondary brain damage and giving the best environment for brain recovery from primary brain injury. Secondary brain damage occurs from hypoxia, hypotension, intracranial haematomas or brain oedema. Prevention, early detection and treatment of the causes of secondary brain damage are the cornerstones of successful head injury management, rehabilitation and good outcome.

Epidemiology of head injuries

Head trauma is a major public health problem worldwide that causes death and disability in young people and makes considerable demands on health resources. Epidemiological studies are essential to initiate appropriate preventive measures and plan necessary services. However, routinely collected data are often unreliable sources of statistics for planning these services. International statistics for accidental deaths and road traffic accident (RTA) deaths do not categorize head injuries as a cause of death independent of other injuries. RTAs are the tenth cause of death throughout the world (2.1% of worldwide mortality) and the incidence of head injuries varies from country to country depending on road and health safeties, e.g. RTAs in the UK are the fourth cause of death (7.7%). International statistics list differences in accident rates between countries and over time. For example, RTA deaths are more frequent in France, Australia and the USA. Head injuries in the UK affect 6–10 per 100 000 of the population and 20% of patients with head injuries attending emergency departments in the UK are admitted to hospital, with 0.2% mortality. Head injuries are steadily decreasing each year in developed countries owing to increased road safety and safety at the work place, e.g. seat belts, better car designs, better roads, speed cameras, and implementation of health and safety regulations by employers. However, head injuries are still responsible for 25–33% of accidental deaths in hospitals and more than 60% of traumatic deaths. To reduce morbidity and mortality of head injuries, several guidelines and protocols have been developed and implemented worldwide. The National Institute for Clinical Excellence (NICE) and the Scottish Intercollegiate Guidelines Network are just a few to list in this chapter.

Head injuries vary from mild concussion to severe brain injury leading to severe disability or death. The successful management of a patient following head trauma requires understanding of the pathophysiological processes that occurred and processes that might occur after a head injury.

Pathophysiology of head injuries

Most head injuries are caused by blunt trauma. In adults such injuries are caused by RTAs, assaults and falls, whereas in children the commonest causes are falls and injuries from recreational activities such as biking, skating and skateboarding. A small but important cause of head trauma in children is non-accidental head injuries. The mechanisms of head trauma include those listed below.

Direct trauma

In blunt trauma, the energy of impact is spread over a large area of the skull, whereas penetrating injuries cause most of their damage by direct trauma.

Cerebral contusion

Contusion of the brain may occur at the site of the impact – 'coup injury' – but it is most common and most severe at the opposite site of the primary impact – 'contrecoup injury' (Figure 15.1). As the brain is stationary within the skull, if the back of the head is struck by an object causing local scalp laceration, bruising, skull fracture or focal brain contusion, the sudden acceleration and deceleration of the brain within the skull causes the opposite poles of the brain to collide against the inside ridges of the skull bone, leading to severe brain contusions, e.g. a sudden blow to the occiput may cause scalp laceration, cephalohaematoma, skull fracture or occipital contusion. In addition, owing to the brain acceleration and deceleration the temporal and frontal lobes slide against the sphenoid ridge and the floor of the anterior and middle cranial fossae, leading to severe contusion of the undersurface of the frontal and temporal lobes (Figure 15.1).

Diffuse axonal injury

Diffuse axonal injury is caused by sheering forces from acceleration and deceleration of the brain following blunt trauma. These rotation forces cause petechial haemorrhages, particularly in the upper brainstem, cerebellum and corpus callosum, that are seen as retraction balls and microglial scars early on, and widespread myelin degeneration a few weeks after the injury. CT scan demonstrates multiple petechial haemorrhages (Figure 15.2).

Cerebral oedema

Cerebral oedema occurs as a result of vasomotor failure leading to brain swelling around the cerebral contusions or as a result of cytotoxic effects throughout the brain. Brain oedema can cause raised intracranial pressure (ICP). Raised ICP can lead to brain herniation from one side of the supratentorial compartment to another (subfalcine), or the medial temporal lobe down the tentorial hiatus compressing the ipsilateral third nerve (true localizing sign) and pushing the brainstem against the contralateral edge of the tentorium, leading to ipsilateral hemiparesis (false localizing sign). This type of transtentorial descent of the temporal lobe is called transtentorial herniation. The final stage of raised ICP causes descent of the cerebellar tonsils in the foramen magnum, leading to brainstem compression and respiratory arrest. At postmortem the impression on the cerebellar tonsils caused by transforaminal magnum herniation is described as coning, and the impression of transtentorial brain herniation on the contralateral brainstem is called a Kernohan notch (Figure 15.3).

Traumatic intracranial haematomas

Traumatic intracranial haematomas include extradural, subdural, subarachnoid and intracerebral haemorrhages. Traumatic intracranial haematomas can cause similar brain shifts to those described under cerebral oedema.

Extradural haematoma

Extradural haematoma is blood clot between the skull bone and the dura. The classic history of an extradural haematoma is that a patient was concussed with brief loss of consciousness followed by a lucid interval that can last minutes to hours during which the patient appears normal, followed by sudden deterioration in the level of consciousness. The classic history occurs in <27% of extradural haematomas. In 85% of patients, the source of bleeding is the middle meningeal artery; in young patients (<30

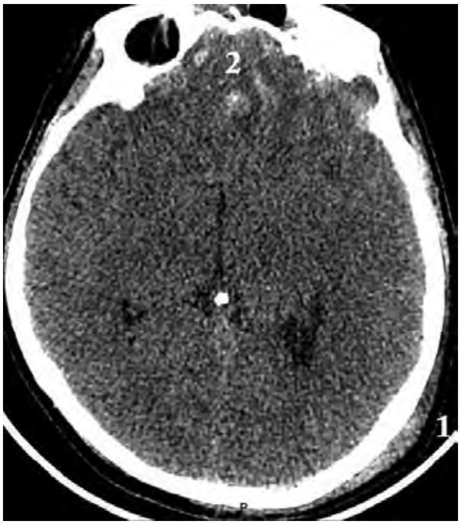


Figure 15.1 Axial CT of the head demonstrating small cephalohaematoma (1) at the impact site and contrecoup contusions (2) at the under-surface of both frontal and left temporal lobes.

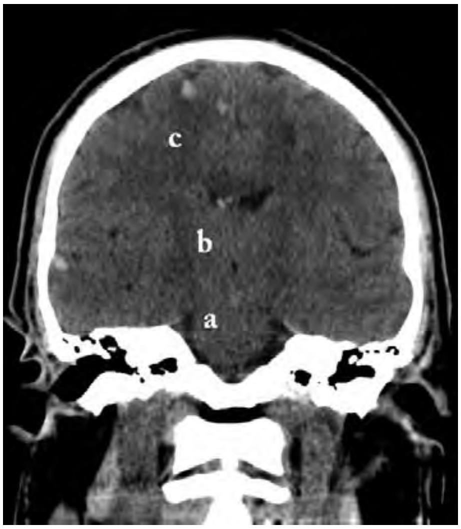


Figure 15.2 Coronal CT of the head demonstrating petechial haemorrhages in the midbrain (a), corpus callosum (b) and right hemisphere (c).

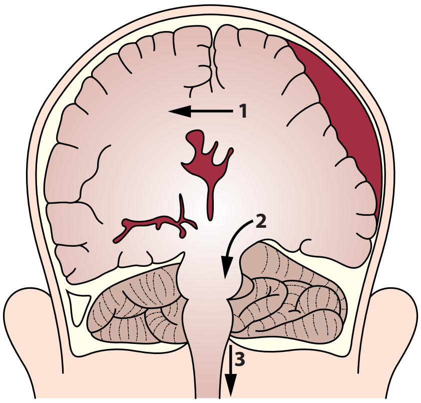


Figure 15.3 The coronal section of the head demonstrating an extradural haematoma (4) compressing the left hemisphere, causing subfalcine brain hernia (1), transtentorial brain hernia (2) and transforaminal magnum tonsillar hernia (3).

years old) 40% of extradural haematomas were associated with fracture of the cranial vault, whereas in those >30 years old skull fracture was found in almost all patients. It is associated with acute subdural haematoma in 20% of patients and the mortality rate from extradural haematoma was 20–55%; however, with new head injury guidelines and early treatment the mortality from extradural haematoma has been reduced to 5–10%.

On CT scan extradural haematoma appears as a high-density collection beneath the skull bone, taking the shape of a bilenticular (biconvex) lens (Figure 15.4). Symptomatic extradural haematoma requires removal via craniotomy.

Subdural haematoma

Subdural haematoma is a collection of blood in the subdural space and used to be classified according to the age of the haematoma into acute if discovered within 72 hours; subacute if discovered within 3 days to 3 weeks; and chronic if the haematoma was >3 weeks old. However, most patients do not remember the onset of head injury and therefore this classification became obsolete after CT and MRI scans were introduced. The most relevant classification clinically is to classify these lesions as follows.

Hyperdense subdural haematoma

This occurs when the clot is higher density than the brain on CT and takes the shape of the brain, as demonstrated in Figure 15.5. If the lesion is symptomatic, its removal requires craniotomy. If it is >1 cm thick it generally needs to be removed. Its volume can be measured by the formula: (thickness × height × length)/2. If the volume is >80 mL, removal is also recommended. Other factors that influence the decision-making process include mass effect in the form of midline shift and the Glasgow Coma Scale (GCS) score. If a hyperdense subdural haematoma is removed within 4 hours of injury, survival can be expected in up to 70% of patients compared with only 10% if the surgical evacuation is delayed >4 hours or if the patient's pupils are unreactive. More than 79% of patients who had a hyperdense subdural haematoma that was evacuated surgically develop raised ICP; therefore, monitoring of ICP is desirable in these patients after surgery.

Mixed density subdural haematoma

This occurs when the density of the clot is a mixture of high density and isodensity in appearance. This is usually treated in the same fashion as hyperdense subdural haematoma.

Isodense subdural haematoma

This is when the density of subdural haematoma is the same as that of the brain. This can be a challenging diagnosis and keeping a high index of suspicion is essential to detect these lesions on CT. They can be detected by observing indirect signs such as midline shift, effacement of the lateral ventricle or effacement of the ipsilateral sulci. Intravenous contrast injection will often show enhancement of the subdural membrane and makes visibility easier. These lesions can be evacuated through two burr holes.

Hypodense subdural haematoma

This is when the collection has a lower density than the brain. This can be drained via one or two burr holes. If the brain surface did not come to the skull surface during drainage, it is advisable to leave a subdural drain *in situ* to drain any residual clot and avoid tension pneumocephalus. The latter can arise if a hypodense subdural haematoma is drained while the patient's head is elevated and the cavity is filled with air at room temperature that will heat up during the next few hours. The air then will expand according to Boyle's law, leading to tension pneumocephalus. Routine scans are not required after drainage of a hypodense subdural haematoma, as removal of 20% of the fluid normalizes the ICP and 78% of patients would have residual collection by day 10 and 15% by day 40 after drainage. The important step to prevent recurrence of symptoms of hypodense subdural haematoma is to leave the dura within the burr hole open and make a small pocket in the subgaleal space for any residual collection to be absorbed.

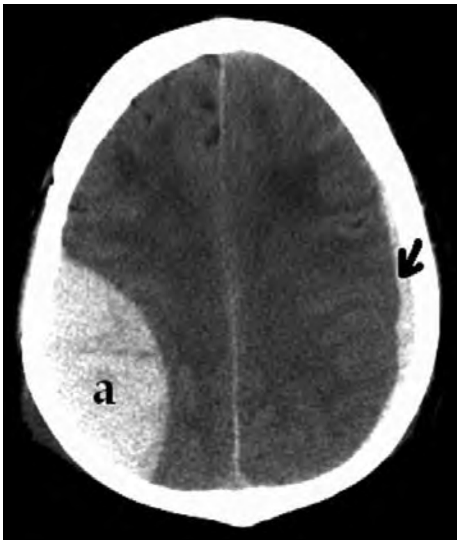


Figure 15.4 Axial CT scan demonstrating a right extradural haematoma (a) and a contralateral acute subdural haematoma (arrow).

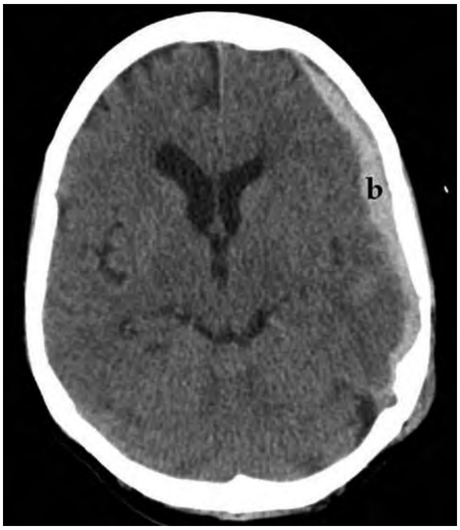


Figure 15.5 Axial CT scan demonstrating hyperdense subdural haematoma (b). The shape of the collection is convex to the outside towards the skull surface and concave towards the brain surface. The hyperdense subdural haematoma caused subfalcine brain herniation.

More than three-quarters of hypodense subdural haematomas are treated by a single burr hole (Figure 15.6).

Traumatic intracerebral haematoma

Traumatic intracerebral haematoma often occurs in addition to acute subdural haematoma or contusion. The difference between cerebral contusion and traumatic intracerebral haematoma is arbitrary. Both are characterized by hyperdense parenchymal lesions on CT scan. Contusions often reach the cerebral cortex and they appear bigger than their actual mass effect, whereas traumatic intracerebral haematomas appear more distinct and their size and mass effect appear proportionate to a midline shift (Figure 15.7).

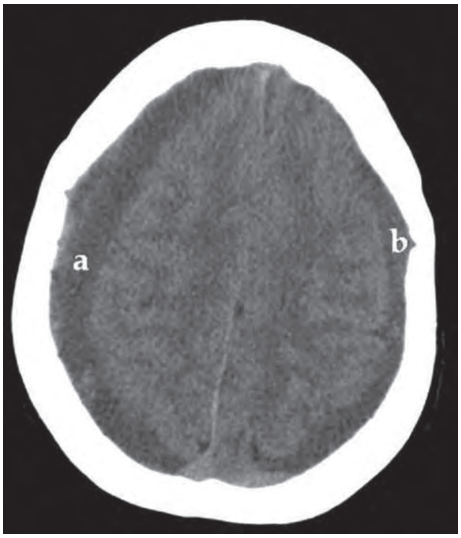


Figure 15.6 Axial CT of the head demonstrating bilateral hypodense subdural collections (a,b). The right is bigger than the left and both are lower density than the brain. These are the CT characteristics of chronic subdural haematoma (hypodense subdural haematoma).

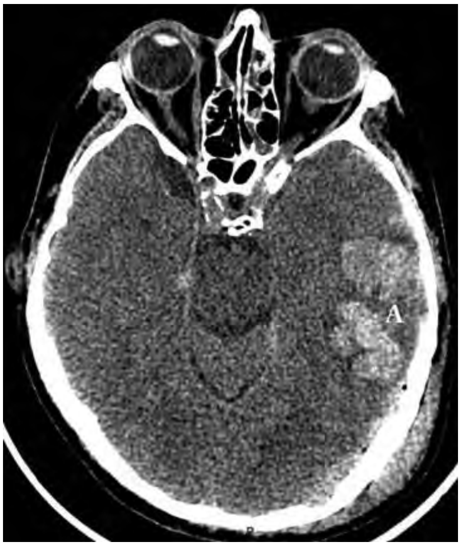


Figure 15.7 Axial CT scan of the head demonstrating a large hyperdense parenchymal lesion in the left temporal lobe (A), a traumatic intracerebral haematoma.

Traumatic subarachnoid haemorrhage

Traumatic subarachnoid haemorrhage is blood in the subarachnoid space following trauma, and frequently occurs in moderate or severe traumatic brain injury and is related to worse outcome at time of discharge. The presence of traumatic subarachnoid haemorrhage appears to be associated with worse vocational outcome in survivors of moderate or severe traumatic brain injury. As such, the presence of traumatic subarachnoid haemorrhage appears to have predictive value with respect to outcome.

Hydrocephalus

Hydrocephalus is an uncommon cause of raised ICP in head injuries in the acute stage. It can happen when a blood clot obstructs the fourth ventricle or the cerebral aqueduct. Hydrocephalus may also follow traumatic subarachnoid haemorrhage because traumatic subarachnoid haemorrhage blocks the cerebrospinal fluid (CSF) absorption sites. More often unilateral hydrocephalus follows subfalcine brain hernia owing to obstruction of the third ventricle. When contralateral hydrocephalus occurs as a result of raised ICP, the raised ICP becomes worse and brain hernia ensues (Figure 15.8).

Concussion

Concussion is transient loss of consciousness following non-penetrating closed head injury without gross or microscopic brain damage. Concussion is divided into three grades as follows:

Grade 1 No loss of consciousness; confusion without amnesia.
Grade 2 No loss of consciousness; confusion with amnesia.
Grade 3 Loss of consciousness.

Other injuries associated with brain injuries

In addition to concussion, cerebral contusion, intracranial haematomas and cerebral oedema, the following injuries may

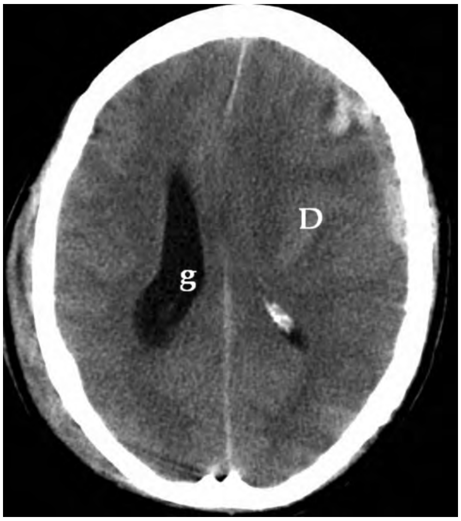


Figure 15.8 Axial CT scan of the head demonstrating very extensive brain swelling (D), causing severe subfalcine brain hernia and contralateral hydrocephalus (g).

occur: scalp lacerations, scalp haematomas, skull fractures, cranial nerve injuries, vascular injuries and spinal injuries. These injuries need to be considered in patients with head trauma.

Scalp lacerations

The extent of scalp lacerations does not indicate the severity of brain injury. However, they need to be treated promptly to stop blood loss and to make sure that there is no associated depressed skull fracture. The scalp has five layers: **S** for skin; **C** for connective tissue where the blood vessels and nerves are located, hence the blood vessels are kept open and lead to excessive blood loss unless the bleeding is stopped by compression; **A** for aponeurosis that keeps the scalp stretched; the aponeurosis must be closed when suturing scalp wounds to prevent stretching of the scar; **L** for loose areolar tissue; and **P** is for pericranium. Scalp lacerations must be cleansed and closed, making sure the aponeurosis is approximated to prevent the edges from retraction and widening the scar. Local anaesthesia, to be effective, must be injected in the connective tissue layer above the aponeurosis.

Cephalohaematomas

Haematomas of the scalp are not uncommon and do not need specific treatment. However, a large cephalohaematoma in an infant may represent significant blood loss. Cephalohaematomas around the eyes are often associated with basilar skull fracture of the anterior cranial fossa and CSF leaks (panda eyes or racoon eyes). Cephalohaematoma around the mastoid is often referred to as Battle's sign and is associated with basilar skull fracture of the petrous bone and otorrhoea.

Skull fracture

Injuries to the head may cause skull fractures. These are classified into simple linear fracture of the skull vault (Figure 15.9); depressed skull fracture if bone fragments were depressed deeper than the inner table of the skull; or compound skull fracture when the skin or the mucosa overlying the skull fracture is deficient, e.g. via scalp laceration or basal skull fracture crossing the paranasal sinuses, the middle ear or the mastoid. Basilar and simple linear skull fractures do not require specific treatment but they could indicate serious underlying brain injury, e.g. basilar skull fractures could be associated with CSF leaks and cranial nerve injuries, i.e. olfactory, optic, vestibulocochlear and facial nerves. Compound skull fractures by definition are contaminated and require wound lavage, primary closure and antibiotic treatment for 7 days according to local antibiotic policy. Depressed skull fractures on the other hand may need treatment if they are compound, depressed more than the thickness of the skull or cosmetically unacceptable except when they overly venous sinuses.

Dural fistulas and cerebrospinal fluid leaks

CSF rhinorrhoea or otorrhoea indicates that there has been a breach of the dura mater. This often occurs in association with skull base fractures involving the anterior cranial fossa, sphenoid or temporal bones. It is often difficult to be sure that blood-stained nasal or ear discharge contains CSF. Glucose and protein estimations in the discharge are often positive, even in the absence of CSF. The best test to confirm the presence of CSF is to measure β_2-transferrin (β_2T). β_2T is present only in the CSF and aqueous humour of the eye. So, if there is any doubt, fluid should be collected and sent for β_2T measurement. The reason why CSF leaks are treated seriously is the high risk of recurrent meningitis in these patients, often due to *Streptococcus pneumoniae*, which still carries high mortality. Prophylactic antibiotics are not indicated here because they merely change the type of organism rather than prevent meningitis. CSF leaks can be diagnosed and localized using fine-slice CT to demonstrate the bony defect and MRI scan to demonstrate CSF fistula (Figure 15.10) or brain hernia through the fistula. Persistent CSF leaks and those complicated by meningitis require dural repair.

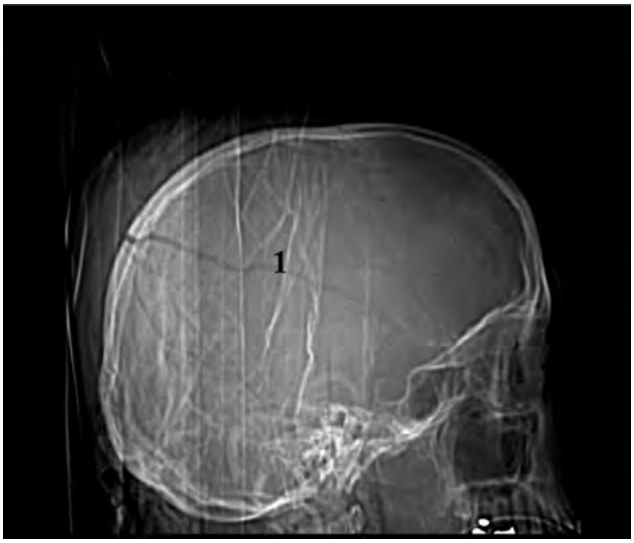


Figure 15.9 Sagittal scanogram demonstrating linear skull fracture (1) in the parietotemporal bone.

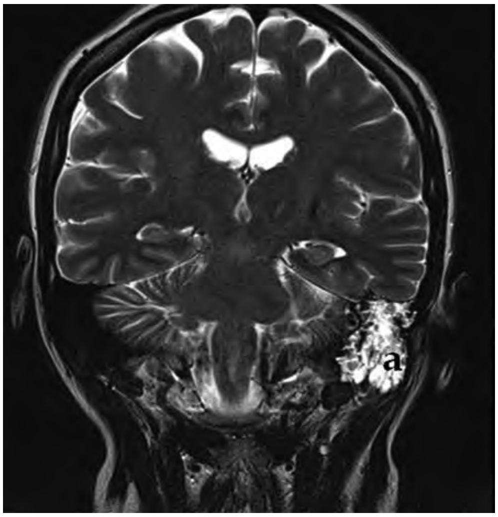


Figure 15.10 Coronal T_2 weighted magnetic resonance image demonstrating a cerebrospinal fluid (CSF) fistula in the left petrous temporal bone with CSF filling the left mastoid air cells (a).

Cranial nerve injuries

Anterior skull basal fractures can be associated with injuries of the first (olfactory) and second (optic) cranial nerves. While the first cranial nerve can be injured and lead to anosmia after minor head injuries, the second cranial nerve is more resistant to injury. Optic nerve injury occurs in about 5% of moderate to severe head injuries. The fourth cranial nerve is susceptible to injury leading to diplopia because of its slender long intracranial course. The sixth nerve is also susceptible to injury because of its long intracranial course. Sixth nerve palsy in head injuries is a false localizing sign. The seventh nerve is susceptible to injury because of its long bony course in the facial canal leading to facial weakness. Facial palsy can occur early owing to direct nerve injury, or late owing to oedema and swelling of the nerve in the facial canal. Injury to the cochlear nerve leads to deafness.

Other injuries

Head injuries do not occur in isolation and it is paramount to look for and treat associated spinal and systemic injuries such as chest, vascular, skeletal and abdominal injuries. A high index of suspicion is essential to discover and treat these injuries. Doctors involved in early management of victims of trauma need to avoid the temptation of labelling a patient as having an isolated head injury unless other injuries are ruled out beyond doubt.

Initial management of head injuries

Initial management of head injuries involves adequate clinical assessment of the head injuries and any associated injuries to make the diagnosis of underlying pathological processes, provide timely appropriate treatment and detect any neurological deterioration. The initial assessment should include the assessment of airways, breathing and circulation, followed by a primary survey looking for life-threatening conditions that require treatment on the spot; these include haemothorax, haemoperitoneum, cardiac tamponade, tension pneumothorax and intracranial haematomas. This must be followed by a complete history and physical examination, paying special attention to consciousness, pupils, limbs and cranial nerves, detecting in the process any lacerations, cephalohaematomas, panda eyes, Battle's sign, CSF leakage, haemotympanum, Le Fort or orbital rim fractures, proptosis or carotid bruit. Once the patient becomes stable he or she should be transferred immediately to an appropriate emergency department. It is essential to obtain the initial neurological state of the patient after the injury from witnesses and rescue service personnel to detect neurological deterioration.

Neurological assessment

Neurological assessment must include assessing the level of consciousness, speech and language, memory, cranial nerve examination, motor and sensory examination, and co-ordination. The time of each neurological assessment must be recorded to monitor neurological progress. Neurological observation must be recorded on the observation chart (Figure 15.11) every 15 minutes four times; if the patient is stable or improving, the observation is continued half hourly four times; if the patient is stable or improving, the observation is continued hourly four times; if the patient is stable or improving, the observation is continued 2 hourly four times; if the patient is stable or improving, the observation is continued 4 hourly for the rest of inhospital stay. Any neurological deterioration should be investigated to detect and treat the underlying pathological processes.

Assessment of level of consciousness

Level of consciousness is assessed at the bedside by the GCS. The GCS consists of observing the patient's responses to verbal or painful stimulation. Three responses are observed: best eye-opening response, best verbal response and best motor response.

Best eye-opening response

There are four possible best eye-opening responses in any patient: 1, no eye opening to any stimulus; 2, eye opening to painful stimuli; 3, eye opening to verbal stimuli; and 4, eye opening spontaneously (Table 15.1). One drawback of the best eye-opening response assessment is that it cannot be assessed in patients who have bilateral complete third nerve palsies or bilateral orbital haematomas. In these patients, the best eye-opening response should be recorded as 'C' for closed eyes rather than '1' for no eye-opening response. In patients who have either of these abnormalities in one eye only, the response of the better eye should be recorded for the purpose of level of consciousness assessment.

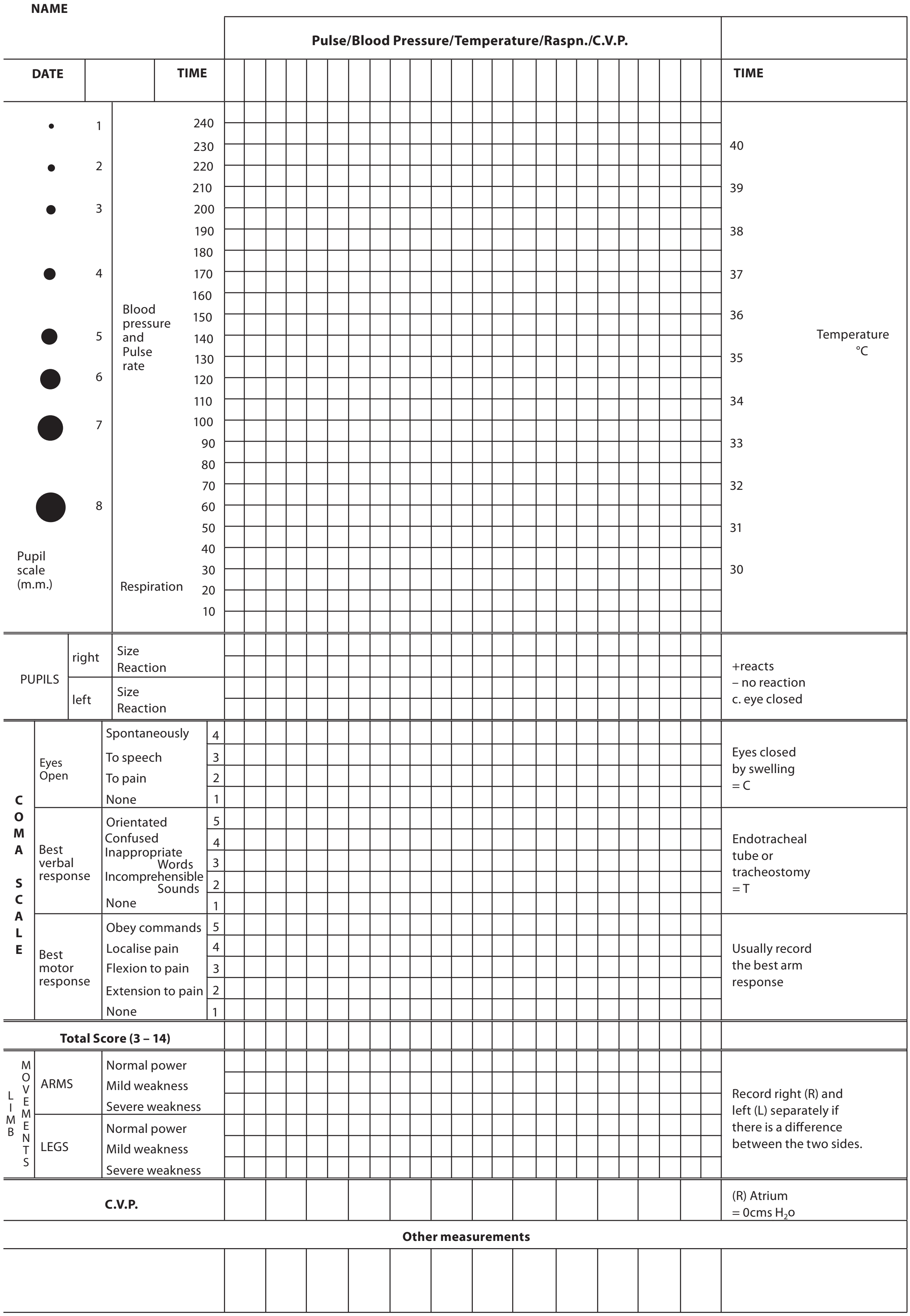
Best verbal response

There are five possible responses under best verbal response: 1, no verbal response to any stimulus; 2, incomprehensible sounds; 3, uttering words; 4, confused; and 5, oriented in time, place and person. It would not be possible to assess best verbal response in patients who are dysphasic and it should be recorded as 'D' for dysphasia rather than '1' for no verbal response. Similarly, patients who are artificially ventilated via endotracheal tube or tracheotomy cannot be assessed for best verbal response and it should be recorded as 'T' for tube rather than '1' for no verbal response (Table 15.1).

Best motor response

There are six possible responses within this category: 1, no motor response to any stimulus; 2, extension to pain; 3, abnormal flexion to pain; 4, flexion to pain; 5, localizing pain; and 6, obeying simple commands (Table 15.1). It would not be easy to assess best motor response in patients with spinal cord injury leading to tetraplegia, although, if they can obey simple commands, best motor response can be assessed by observing motor responses in the face area, e.g. closing the eyes or showing the tongue. If the best motor response cannot be assessed because of paralysis of the limbs due to injury or because of sedative or muscle relaxant drugs, the best motor response should be recorded as 'P' for paralysis rather than '1' for no motor response.

Coma is defined on the GCS as any patient who fulfils all of the following three criteria: 1, no eye-opening response to



NAME

DATE		TIME	Pulse/Blood Pressure/Temperature/Raspn./C.V.P.	TIME
●	1	Blood pressure and Pulse rate	240	
			230	40
●	2		220	
			210	39
●	3		200	
			190	38
			180	
●	4		170	37
			160	
			150	36
●	5		140	Temperature °C
			130	35
●	6		120	
			110	34
●	7		100	
			90	33
			80	
			70	32
●	8		60	
			50	31
			40	
Pupil scale (m.m.)		Respiration	30	30
			20	
			10	

PUPILS	right	Size Reaction		+reacts – no reaction c. eye closed
	left	Size Reaction		
COMA SCALE	Eyes Open	Spontaneously	4	Eyes closed by swelling = C
		To speech	3	
		To pain	2	
		None	1	
	Best verbal response	Orientated	5	Endotracheal tube or tracheostomy = T
		Confused	4	
		Inappropriate Words	3	
		Incomprehensible Sounds	2	
		None	1	
	Best motor response	Obey commands	5	Usually record the best arm response
		Localise pain	4	
		Flexion to pain	3	
		Extension to pain	2	
		None	1	
Total Score (3 – 14)				
LIMB MOVEMENTS	ARMS	Normal power		Record right (R) and left (L) separately if there is a difference between the two sides.
		Mild weakness		
		Severe weakness		
	LEGS	Normal power		
		Mild weakness		
		Severe weakness		
C.V.P.				(R) Atrium = 0cms H_2o

Other measurements

Figure 15.11 Neurological observations.

Table 15.1 The Glasgow Coma Scale (GCS)

Score	Best eye-opening response	Best verbal response	Best motor response	Definition
1	None	None	None	Any patient within these shaded areas is in coma
2	To speech	Sounds	Extension to pain	
3	To pain	Words	Abnormal flexion to pain	
4	Spontaneous	Confused	Flexion to pain	
5		Orientated	Localizing pain	
6			Obeys simple commands	Not in coma

any stimulus (score of 1); 2, no comprehensible sound (score of 1 or 2); and 3, not obeying simple commands (score of 1–5). Therefore, coma is a GCS score of ≤8 provided that the patient does not obey commands, does not utter any words and does not open his or her eyes to any stimulus.

Although each response of the GCS carries a number against it, the aggregated numbers should not be used to describe patients as that leads to avoidable misunderstandings and misinterpretations of the numbers. For example, a GCS of 8 could mean a patient with a best eye-opening response of 1, a best verbal response of 2 and a best motor response of 5 (patient is comatose) or it could mean a best eye-opening response of 2, a best verbal response of 3 and a best motor response of 3 (the patient is not comatose). It would be far better to use the actual description of the patient's response in each category rather than the numbers. Aggregate numbers are useful in determining trends of progress and to perform statistical analysis in research and audit studies.

Categorization of head injuries

Once the initial assessment of a head-injured patient is performed, the risk of raised ICP and brain injury can be categorized as follows:

- *Low-risk head injury*: asymptomatic – there was no loss of consciousness, the patient is fully conscious and oriented and there is no clinical or radiological evidence of skull fracture. The risk of significant intracranial pathology in this group is <9/10 000. The skull radiograph is normal in 99.6% and patients can be discharged home provided there is a responsible adult who can observe the patient and has been given written head injury advice.
- *Moderate-risk head injury*: if there is a history of loss of consciousness, progressive headache or vomiting, post-traumatic seizures, post-traumatic amnesia, an unreliable history or there are signs of basilar fracture or penetrating facial injury. These patients can only be discharged home if they fulfil all the following criteria: normal CT of the brain, initial GCS score >13, discharge GCS score of 15, there is a responsible adult who can look after them and bring them back to hospital if necessary, there are no complicating factors such as non-accidental injury or domestic violence and the patient is given written head injury advice.
- *High-risk head injury*: any patients with a reduced level of consciousness, neurological deterioration, focal neurological deficit, penetrating head injury or depressed skull fracture are a high risk for brain injury.

Management of head injuries

Patients with a GCS score >13 should be admitted to hospital with the following instructions:

1. Activity level: bed rest with head elevated by 30–45°.
2. Neurological observation every 2 hours; if concerned, hourly.
3. Feeding: nil by mouth until fully alert; advance as tolerated.
4. Analgesia: simple analgesics such as tramadol or codeine.
5. Antiemetics: as required.

Patients with a GCS score of 9–13 should also be admitted to hospital with the following instructions:

1. Activity level: bed rest with head elevated by 30–45°.
2. Neurological observations every hour; consider high-dependency unit.
3. Feeding: nil by mouth until fully alert; advance as tolerated.
4. Analgesia: simple analgesics such as tramadol or codeine.
5. Antiemetics: as required.
6. Fluids i.v.: 100 mL/h normal saline with 20 mmol/L KCl.
7. If not awake by 12 hours, repeat the CT brain scan.

Patients with a GCS score <9 should be transferred to a dedicated neurosurgical facility. Prior to transfer, consider the following problems:

- If there is any concern about hypoxia or airway protection during interhospital transfer, intubate and artificially ventilate the patient.
- If the patient has had any seizures administer anticonvulsants.
- If there are any concerns about neck injuries, immobilize the spine.
- If there are any concerns about raised ICP, administer mannitol 2 g/kg of 20% and keep the $P\mathrm{aCO}_2$ between 3.5 and 4 kPa during transfer.

When should a head-injured patient be intubated and artificially ventilated?

This should always be performed for the following reasons:

- to protect the airway if the GCS score is <8 or there is a maxillofacial injury
- if there is evidence of raised ICP, e.g. pupillary dilatation, asymmetric pupillary reaction, decerebrate/decorticate posture or progressive neurological deterioration
- to carry out safe interhospital transfer
- to properly assess a combative or agitated patient by CT.

When should a patient be referred to hospital after a head injury?

When referring a patient to a hospital emergency department, a full history and full neurological examination are required. The history should include the circumstances and timing of the injury, whether the patient lost consciousness, what was the last thing the patient remembers before the injury and what is the first thing he or she remembers after the injury. Amnesia prior to the injury is called pretraumatic or retrograde amnesia. Amnesia after head injury is called post-traumatic or antegrade amnesia. A history of persistent headaches; nausea and vomiting; weakness; sensory disturbance; loss of vision, hearing, sense of smell or balance; diplopia; seizures; or speech impairment is important to note. Clinical examination should include assessment of the level of consciousness, short-term memory, speech and language, vision, hearing, motor power, sensory function and co-ordination. It is also important to note any external bruises on the face and the scalp. Bruising around the mastoid is called Battle's sign and indicates a temporal skull base fracture and may be associated with blood in the middle ear, conductive hearing loss, facial palsy or otorrhoea (CSF leak through the ear) or even paradoxical rhinorrhoea (CSF leak through the nose; this occurs when the tympanic membrane is intact and CSF flows through the fracture into the middle ear, via the Eustachian tube into the nostril). Bilateral black eyes or panda eyes indicate anterior cranial fossa basal skull fracture and may be associated with CSF rhinorrhoea. Scalp lacerations may overlie a depressed skull fracture that needs wound cleaning and fracture elevation. The level of consciousness is assessed by the GCS score (Table 15.1).

A patient who has a head injury should be referred to an emergency department using an emergency ambulance service if:

1. The person presents with an altered level of consciousness. Any person who is not fully conscious at the time of examination after head injury must be referred to the emergency department as an emergency.
2. The patient has had a seizure since the head injury.
3. There is evidence of a basal skull fracture, such as CSF leak from the nose or ear; a black eye with no damage around the eyes (panda eyes); bleeding from one or both ears; haemotympanum (blood behind the ear drum); or bruising around both or one mastoid (Battle's sign).
4. There is evidence of focal neurological deficit, such as double vision, speech impairment, problems with balance, loss of muscle power; or sensory disturbance of the extremities.
5. The patient has a deteriorating level of consciousness.
6. There are associated other injuries such as chest or abdominal trauma; limb or pelvis trauma; or there is significant vascular injury.
7. If the patient does not have access to adequate transportation to hospital.
8. The patient has amnesia for the incident or subsequent events.
9. The patient has any neurological symptoms, e.g. severe and persistent headache, nausea and vomiting, irritability or altered behaviour, seizure, speech impairment, limb weakness or double vision.
10. The mechanism of injury suggests: high-energy injury (e.g. road traffic accident, fall from height), possible penetrating brain injury, possible non-accidental injury (in a child) or continuing uncertainty about the diagnosis after first assessment.
11. The patient has a medical comorbidity, e.g. anticoagulant treatment (warfarin), antiplatelet therapy (aspirin, clopidogrel, dipyridamole), drug abuse or alcohol abuse.
12. There are adverse social factors, e.g. no one able to supervise the patient at home.

When can a person be sent home after minor head injury?

Patients can be sent home if the clinical history and examination indicate a low risk of brain injury and the hospital referral criteria were not met, providing the patient has appropriate support structures and competent supervision at home. Patients must be given verbal and written head injury advice.

When are skull radiographs indicated after head injury?

Skull radiographs should be performed if any of the following apply and if CT is not being performed:

- if the patient is fully conscious and the mechanism of injury has not been trivial, or
- consciousness has been lost, or
- the patient has amnesia or has vomited, or
- the scalp has a full thickness laceration or a boggy haematoma, or
- the history is inadequate, or
- if the level of consciousness is impaired.

When does a CT scan of the head become indicated?

CT scanning should be done in a patient who has any of the following features:

- impaired consciousness
- a deteriorating level of consciousness or progressive focal neurological signs
- confusion or drowsiness followed by failure to improve within at most 4 hours of clinical observation
- radiological or clinical evidence of a fracture, whatever the level of consciousness
- new focal neurological signs that are not getting worse
- full consciousness with no fracture but had other features, e.g.
  - severe and persistent headache
  - nausea and vomiting
  - irritability or altered behaviour
  - development of a seizure.

When do radiographs of the cervical spine become essential?

CT scanning of the cervical spine down to C2 should be undertaken routinely, at the time of head scanning. Imaging of the rest of the cervical spine, including the cervicothoracic junction, should be carried out in the following circumstances:

- fully conscious patient if clinical symptoms or signs or the mechanism of injury indicate the possibility of injury to the spine
- persisting impaired consciousness
- unconscious patient, not localizing pain.

Which patients should be discussed with a neurosurgeon after head injury?

A head-injured patient should be discussed with a neurosurgeon if one or more of the following features existed:

- a CT scan in a general hospital showed a recent intracranial lesion
- the criteria for CT scanning are fulfilled but CT cannot be done within an appropriate time frame locally
- irrespective of the result of any CT scan, if the patient fulfils any of the following criteria:
  - persisting coma after initial resuscitation
  - persisting confusion for more than 4 hours
  - deterioration in the level of consciousness after admission (a sustained drop of one point on the motor or verbal subscales, or two points on the eye-opening subscale of the GCS)
  - progressive focal neurological signs
  - seizure without full recovery
  - compound depressed skull fracture
  - definite or suspected penetrating injury
  - CSF leak or other sign of a basal skull fracture.

Raised intracranial pressure and intracranial pressure monitoring

Raised ICP in head injuries is caused by cerebral oedema or mass lesions such cerebral contusion, intracerebral haematoma, subdural haematoma or extradural haematoma. Treatment of raised ICP requires the treatment of the underlying cause. Significant intracranial haematomas are evacuated and residual raised ICP should be monitored closely to see the response to therapy. The following are some of the indications for ICP monitoring:

- severe head injury: GCS score <8.
- following evacuation of a mass lesion
- ventilated patients with an abnormal CT
- multiple injuries.

ICP monitoring should be avoided in awake patients and in the presence of coagulopathy. ICP monitoring is not without risks: the risk of infection is approximately 1–2%, the risk of bleeding is 2.8% and the risk of malfunction is 6–40%. ICP monitoring can be performed by inserting a transducer in the subdural space, ventricle or brain parenchyma, e.g. a Camino bolt, Codman transducer or Spielberg transducer. Most neurosurgical units treat any ICP ≥25 mmHg to maintain cerebral perfusion pressure of >70 mmHg. Cerebral perfusion pressure is measured by subtracting the mean ICP from the mean arterial blood pressure. Causes of raised ICP include reduced venous return, e.g. severe neck flexion or head-down position, venous thrombosis, sustained seizures or intracranial pathology, such as extradural haematoma, subdural haematoma, cerebral oedema, diffuse axonal injury or ischaemia.

What is the relationship between intracranial pressure and cerebral perfusion pressure?

The cerebral perfusion pressure remains constant when the mean arterial blood pressure is 60–160 mmHg in normal individuals; however, in head-injured patients and those with subarachnoid haemorrhage autoregulation is lost and the cerebral perfusion pressure is directly related to mean arterial blood pressure (Figure 15.12).

How to treat raised intracranial pressure

Raised intracranial pressure should be treated initially with simple measures such as adjusting the head and neck posture, checking the ventilator settings and controlling any seizures or pyrexia.

If simple head and neck position changes do not normalize ICP, the second-tier manoeuvres are instituted: increasing mean arterial blood pressure, for example, leads to increased cerebral perfusion pressure, and reducing body temperature to 32–35°C reduces cerebral metabolism and reduces ICP in the process; however, the patient should be monitored for reduced cardiac output, pancreatitis and renal impairment.

Hyperventilation and reducing $P\text{aCO}_2$ to 3 kPa in the short term will lead to vasoconstriction that reduces intracranial blood volume and reduced ICP. However, if hyperventilation is maintained for more than a few days, cerebral ischaemia will occur. Surgical decompression of the skull vault reduces ICP; however, the key is to include subtemporal decompression for it to work.

Finally, induced barbiturate coma reduces cerebral metabolism and the ICP; however, this can lead to hypotension.

Mannitol is used to reduce ICP and impart cerebral protection; the dose is 2 g/kg of 20% i.v. over 20 minutes; thereafter, the smallest effective dose can be used. Mannitol is used before CT if there is clinical evidence of raised ICP or mass lesion, or in sudden neurological deterioration. Mannitol is also used to assess salvageability in patients with severe head injuries with unreactive pupils. After CT, mannitol is used in patients with mass lesions with raised ICP just before and during surgery. Mannitol would not work during hypotension and raised plasma osmolality (>300 mosmol/L).

The role of exploratory burr holes is now questionable because of the wider availability of CT scan and safe, speedy interhospital transfer of patients. However, rarely this may have to be used when a patient suddenly deteriorates with pupillary dilatation after a lucid interval when it is not practical or possible for the patient to undergo CT. A burr hole can then be made in

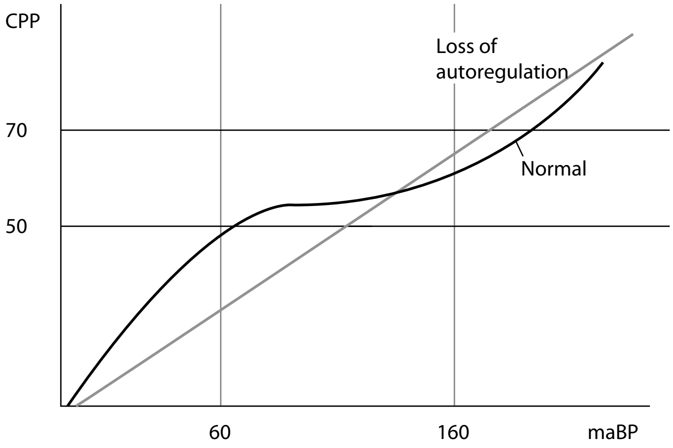


Figure 15.12 The relationship between cerebral perfusion pressure (CPP) and mean arterial blood pressure (maBP) in normal and abnormal autoregulation.

the temporal area 2 cm in front and above the tragus on the side of the first pupillary dilatation; then a frontal burr hole should be made just behind the hairline at least two finger breadths from the midline; finally, a parietal burr hole should be formed over the parietal eminence.

Advice for patients after head injury on discharge home

Patients should be instructed that if they are affected by any of the following after leaving the hospital, they should return to their nearest hospital emergency department as soon as possible:

- unconsciousness, or lack of full consciousness (e.g. problems keeping their eyes open)
- any confusion (not knowing where they are, getting things muddled up)
- any drowsiness (feeling sleepy) that goes on for longer than 1 hour when they would normally be wide awake
- any problems understanding or speaking
- any loss of balance or problems walking
- any weakness in one or both arms or legs
- any problems with their eyesight
- a very painful headache that will not go away
- any vomiting
- any fits (collapsing or passing out suddenly)
- clear fluid coming out of their ear or nose
- bleeding from one or both ears
- new deafness in one or both ears.

Things patients should not worry about

It is important that patients understand that, over the next few days, they may experience one or more symptoms, which usually disappear in the following 2 weeks. These symptoms include a mild headache, feeling sick (without vomiting), dizziness, irritability or bad temper, problems concentrating or problems with their short-term memory, tiredness, lack of appetite or problems sleeping. If patients are very concerned about any of these symptoms in the first few days after discharge, they should go and see their own doctor.

Finally, patients should be told that, if these problems do not go away after 2 weeks, they should go and see their doctor anyway. Ideally, patients should also seek a doctor's opinion about their ability to drive a car or a motorbike.

GUIDE TO FURTHER READING

Eljamel MS. Antibiotic prophylaxis in CSF fistulae. *Br J Neurosurg* 1993;**7**:501–506.

Eljamel MS, Foy PM. Acute cerebrospinal fluid fistulae, the risk of intracranial infections. *Br J Neurosurg* 1990;**4**:381–385.

Eljamel MS, Waring DJ. The Paragon immunofixation for CSF identification. *Biomed Sci* 1993;**4**:43–45.

Eljamel MS, Pidgeon CN, Toland J, *et al*. MRI-cisternography in CSF fistulae localisation. *Br J Neurosurg* 1994;**8**:433–437.

Eljamel S (ed.). *Problem Based Neurosurgery: Clinical Examination.* Singapore: World Scientific Publishing, 2010:12.

Griffiths C, Brock A. Twentieth century mortality trends in England and Wales. *Health Stat Q* 2003;**18**:5–17.

Hanlon RE, Demery JA, Kuczen C, Kelly JP. Effect of traumatic subarachnoid haemorrhage on neuropsychological profiles and vocational outcome following moderate or severe traumatic brain injury. *Brain Injury* 2005;**19**:257–262.

National Institute for Health and Clinical Excellence. *Head Injury: Triage, Assessment, Investigation and Early Management of Head Injury in Infants, Children and Adults. Clinical* Guidance 56. London, UK: NICE, 2007. Available from: http://www.nice.org.uk/guidance/index.jsp?action=download&o=36260 http://www.nice.org.uk/Guidance/CG56

Scottish Intercollegiate Guidelines Network. *Early Management of Patients with a Head Injury*. Publication no. 46. Edinburgh, UK: SIGN, 2000. Available from: http://www.sign.ac.uk/guidelines/fulltext/46/

PART II

Specialist general surgery

CHAPTER 16

Disorders of the skin and soft tissues

SIR ALFRED CUSCHIERI

Anatomy and function of the skin

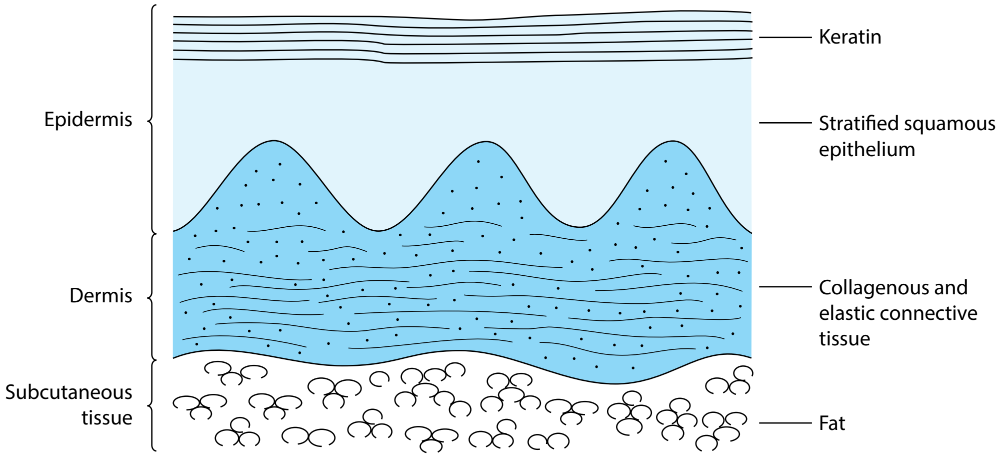
The skin is the largest organ in the body and carries a range of important physiological functions/adaptations, including (1) waterproofing, (2) forming a barrier against infection, (3) vasoconstriction during erect posture and (4) acting as an efficient heat exchanger with the environment in the maintenance of body temperature.

The skin is divided into two main layers: the surface epithelium or epidermis and the underlying dermis (Figure 16.1). The epidermis is a keratinized stratified squamous epithelium and, for descriptive purposes, it is divided into five layers (Figure 16.2) from below upwards:

1. stratum basale
2. stratum spinosum
3. stratum granulosum
4. stratum lucidum
5. stratum corneum.

The stratum basale is the deepest layer and consists of a single layer of cylindrical cells with deeply basophilic cytoplasm and centrally placed nuclei. Interspersed between the basal cells are *melanocytes*, a specialized group of dendritic cells which synthesize the pigment melanin. Above the stratum basale is the stratum spinosum, which consists of two to six rows of cuboidal cells that become flattened towards the surface. Mitotic activity in the resting epidermis occurs mainly in the stratum basale, although, in regeneration following injury, this extends to the stratum spinosum.

The next layer is the stratum granulosum, which is made up of one to three layers of diamond-shaped cells with darkly staining



Epidermis
Dermis
Subcutaneous tissue
Keratin
Stratified squamous epithelium
Collagenous and elastic connective tissue
Fat

Figure 16.1 The anatomy of human skin.

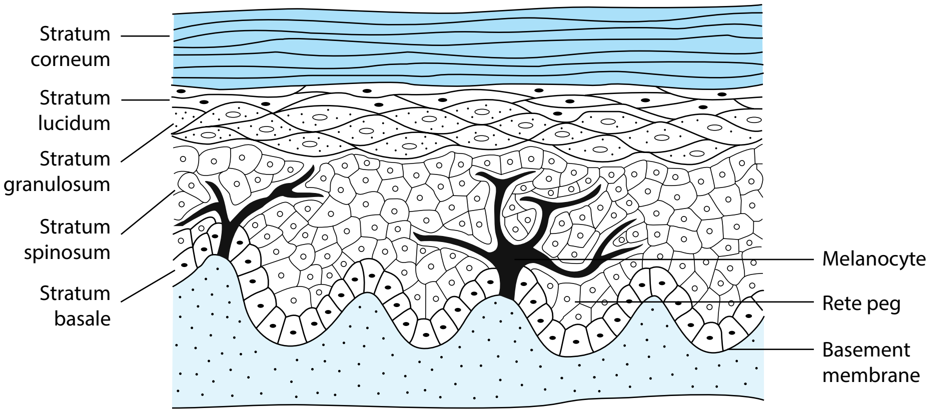


Figure 16.2 The different layers of the epidermis.

pyknotic nuclei. This layer is most obvious in the thick skin of the palms and soles of the feet. This is the layer that produces the fibrous protein *keratin*. In thick skin immediately above the stratum granulosum is a single layer of hyaline anucleate cells termed the stratum lucidum, which is only present in the skin of the palm of the hands and soles of the feet. The outermost layer of the skin is the stratum corneum, consisting of a variable number of flattened anucleate dead cells containing keratin. Over pressure areas such as the hands and the feet, the keratin layer is dense and compact, whereas, over the rest of the body surface, it forms a looser covering.

The under surface of the epidermis consists of downward epithelial projections into the dermis called *rete* or *interpapillary pegs*. The dermis itself is composed of fibroelastic tissue, which supports the epidermis and its appendages. It has two layers: the superficial zone containing the papillary pegs and the deeper reticular layer. The dermis becomes continuous with the denser subcutaneous fascia.

The epidermal appendages

Sweat glands

Sweat glands are found all over the skin and consist of simple tubular glands extending from the epidermis to the mid-dermis, where they become coiled. This is the part of the gland that produces sweat, which is a clear hypotonic electrolyte fluid (with respect to plasma). The main solute of sweat is sodium chloride, but it also contains potassium, lactic acid and urea. Sweating is a mechanism involved in the regulation of body temperature and is controlled by the hypothalamus via cholinergic fibres of the sympathetic nervous system; sweating is stimulated by a rise in body temperature brought about by an increase in environmental temperature or exercise. It is also induced by hypotension, anxiety and fear.

Apocrine glands

These glands are found in the axilla, anogenital region, the mammary areola and the canal of the external ear. Similar to the sweat glands, apocrine glands form coiled tubules which are, however, much larger and situated lower in the dermis, extending into the subjacent subcutaneous fat. Unlike the sweat glands, which penetrate the epidermis, the apocrine glands open into hair follicles above the ducts of the sebaceous glands. The secretion from the apocrine glands has a milky appearance and its production is stimulated by adrenergic stimuli including fear or pain. Apocrine glands become active only after puberty, although their physiological role is not clear. Blockage of the apocrine glands leads to the recurrent inflammatory condition of *hidradenitis suppurativa*.

Hair

Hair grows from hair follicles, which are tubular invaginations of the epidermis into which open sebaceous and apocrine glands. The fact that the epidermis is continuous with the hair follicles is of surgical importance as the follicular epithelium and, to a lesser extent, the epithelium of the sweat glands can grow and migrate to the surface to regenerate the skin after the epithelium has been destroyed by trauma or burns or following procurement of split skin grafts. Each single hair is basically a long, dead keratinized shaft which starts as a small growth area in the depths of the follicle known as the hair bulb. This area is situated in the upper layer of the subcutaneous fat and is invaginated on the underside by vascular connective tissue known as the hair papilla. Above the hair bulb, the cells become elongated and undergo keratinization and, from this point onwards, the hair shaft is a dead structure made up of an outer cortex and a looser inner medulla. Inserted into the walls of the hair follicles are small strips of smooth muscle, the *erectors pilorum*. These are supplied by adrenergic nerve fibres and cause erection of hairs induced by cold and emotional stress.

Sebaceous glands

These glands are absent from the palms of the hands and soles of the feet, but are otherwise found all over the skin surface. They are holocrine glands that produce their secretion by disintegration of the contents of the cells, which are then discharged directly into the sebaceous duct, which enters the hair follicle (Figure 16.3). The function of the secretion (sebum) is to lubricate the skin and act as a physical protective barrier. Sebum consists of a mixture of fatty acids, glycerides, cholesterol and other substances. Sebaceous gland activity rapidly increases after puberty.

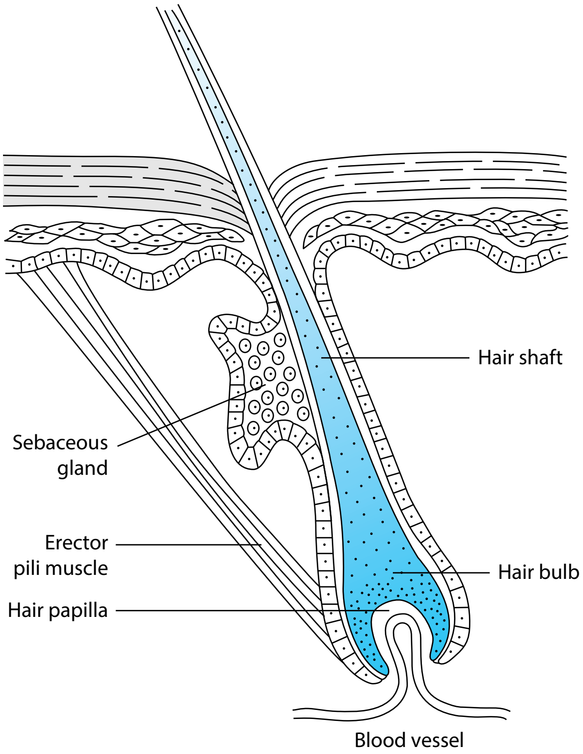


Figure 16.3 Anatomy of the hair follicle.

Nails

Nails are made of semitransparent keratin and are surrounded proximally and laterally by folds of skin. They arise from the nail bed, which consists of a modified epidermis lacking a stratum granulosum. It has a highly vascular dermis, which is continuous on the deep aspect with the periosteum of the distal phalanx.

Pigmentation

In man the main function of pigmentation of the skin appears to be the provision of protection against the harmful effects of sunlight, and dark skin appears to be the result of selective adaptation to intense prolonged exposure to sunlight. The cells responsible for pigmentation are the *melanocytes* found between cells of the stratum basale at the junction of the epidermis and the dermis. These cells synthesize the black pigment *melanin* from the amino acid tyrosine by means of the copper-dependent enzyme oxidase tyrosinase. Melanin is transferred to the surrounding basal cells by the dendritic processes of the melanocytes. Pigmentation by melanin is controlled by the melanocyte-stimulating hormone (MSH) produced by the anterior lobe of the pituitary gland. Excess MSH secretion results in increased melanin pigmentation. Interestingly, the number of melanocytes per unit area of the skin is the same in all races and increased pigmentation therefore reflects an increased melanin synthesis.

Blood supply of skin

The blood supply to the skin not only supplies oxygen and nutrients but also plays an important role in temperature regulation and maintenance of the peripheral vascular resistance during prolonged standing. Skin requires a minimum flow of 0.8 mL of blood per minute per 100 mL of tissue to supply its oxygen requirements, but there is an abundant blood supply which allows considerable variation in blood flow. Indeed, it has been estimated that, during whole body heating, maximal skin blood flow may reach 7–8 L/min. Only skeletal muscle has the potential to receive a greater blood flow.

Small arterial branches penetrate the subcutaneous fat and form a plexus from which loops of vessels go up to supply the individual papillae. Branches from these plexuses supply the glands and the hair roots. The blood is returned via venous plexuses which are situated below the arterial one, and many arteriovenous shunts, known as glomus bodies, are found in certain regions, particularly the palms, the soles, the ears and the central part of the face. These allow the capillary circulation to be short circuited and are concerned with temperature regulation. The very smallest cutaneous arterioles contain muscle fibres, which are again important for thermoregulation. The epidermis itself is avascular and receives nourishment from the vessels at the tip of the papillae via the intercellular spaces. The dermis also has a rich network of lymphatics which start in the papillae and eventually join the larger subcutaneous lymph channels.

In the skin of the trunk and limbs, cutaneous blood flow is controlled by the noradrenergic vasoconstrictor system and an active vasodilator system, in addition to local factors. The skin blood flow response to rapid local heating is biphasic with an initial rapid increase in skin blood flow, followed by a more prolonged plateau phase. The initial phase appears to be largely mediated by an axon reflex, whereas the secondary prolonged hyperaemic phase is caused predominantly by the endothelial release of nitric oxide. The skin hyperaemic response to local heating has caused considerable interest in recent years following the documentation that it is impaired in both type 1 and type 2 diabetic patients and that this reduced capacity for maximal flow can be detected before the onset of diabetes. It is a reflection of impaired endothelial function known to be a feature of this disease.

Nerve supply of the skin

The skin is supplied by cutaneous nerves (eight cervical, 12 thoracic, five lumbar and five sacral) and each area of skin supplied by a single spinal nerve (which form the various named cutaneous nerves) is known as a dermatome (Figure 16.4). The sensory nerve endings are of three forms: (1) endings with expanded tips or discs (Merkel's discs); (2) encapsulated nerve endings (Meissner's corpuscles in the dermal papillae), Ruffini, Pacinian and Golgi–Mazzini corpuscles in the dermis and subcutaneous tissue; and (3) bare nerve endings, which are found in all layers of the skin including base of hair follicles. In addition, recently, it has been demonstrated that some nerve fibres on the cutaneous arterioles may also provide sensory information.

The five primary forms of cutaneous sensation are touch–pressure (mechanoreceptors), cold and warmth (thermoreceptors), pain and itch. Each quality is served by a specific set of cutaneous

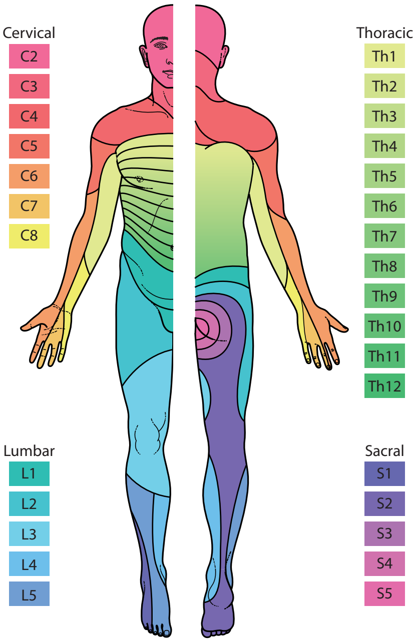


Figure 16.4 Dermatomes of the human skin.

peripheral myelinated and non-myelinated nerve fibres. More complex sensations result from an integration within the central nervous system. The two receptors which are responsible for light touch are Merkel's discs and Meissner's corpuscles. The Pacinian corpuscle is sensitive to pressure. Pruritus (itch) is a common symptom and may be produced by a variety of physical stimuli and disease processes. It seems to be closely related to pain as it travels in the C-fibres of sensory nerves and ascends in the spinothalamic tracts of the spinal cord in association with the pain fibres.

Benign cutaneous surgical conditions

Epidermoid cysts

Often called sebaceous cysts, these lesions are derived from the epidermis lining the hair follicle and sebaceous glands. They may occur anywhere on the skin but particularly on the head, face, neck and trunk. They are unilocular dermal cysts giving rise to visible and palpable cystic swellings which are fixed to the skin (Figure 16.5). There is usually a visible central punctum at the site of the follicle. Histologically, the cysts are made up of layers of epidermal cells interspersed with keratin layers. An epidermoid cyst may become inflamed and present with a tender indurated swelling.

Treatment involves excision along with an ellipse of overlying skin. It is important to excise the whole of the cyst to prevent recurrence. In the infected state, an epidermoid cyst should be incised to allow drainage and the remnant is excised at a subsequent stage when the inflammation has subsided.

True sebaceous cysts occur in a rare condition known as *steatocystoma multiplex*, which is characterized by the presence of multiple cysts containing oily sebum. Another type of cyst which may be mistaken for an epidermoid cyst arises from the epidermis of the external root sheath of the hair follicle. These are known as *pilar cysts* or *trichilemmal cysts*. They tend to occur in women and appear to be familial. Occasionally, the walls may rupture and the resulting proliferative reaction may mimic a squamous cell carcinoma (SCC).

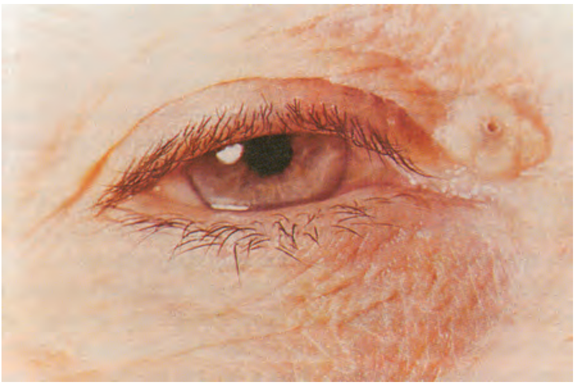


Figure 16.5 Typical epidermoid (commonly known as sebaceous) cyst, this example being situated at the inner canthus of the right eye. (Courtesy of Dr Sally Ibbotson, University of Dundee, UK.)

Molluscum contagiosum

This lesion is caused by the molluscum contagiosum virus, which is a member of the pox virus family. It is contagious but with low virulence in immune-competent patients. It tends to occur in children and young adults, in whom it is often sexually transmitted. Transmission may be direct (skin contact) or through fomites (infected towels and clothing). The incubation period averages 2–3 months. The virus multiplies in the deeper layers of the epidermis and destroys the low-lying tissue producing a central pore. The most typical clinical presentation is a crop of smooth firm, raised pink umbilicated lesions up to 1.0 cm in diameter with a central pus-like head. In adults, the lesions tend to be located on the thighs, buttocks, groin and lower abdomen and even on the genitalia, whereas in children they develop on the face, trunk, legs and arms. The usual form of treatment is curettage but cryotherapy or topical trichloroacetic acid or iodine or phenol can be effective. However, the lesions often resolve without treatment. In HIV patients, however, molluscum contagiosum often persists and takes an aggressive course.

Warts

Warts are exceedingly common, and it is estimated that about 10% of all individuals are affected, usually teenagers, but they may occur at any age. They originate as a result of epithelial hyperplasia triggered by a human papillomavirus (HPV), a group which consists of more than 100 related viruses. Some HPVs types are associated with the development of certain cancers and these are designated as oncogenic or carcinogenic HPVs, e.g. cervical, anal, oropharyngeal and penile cancers. The HPV causing skin warts is not oncogenic. Skin warts are of various types:

- common wart (*verruca vulgaris*): raised nodule with roughened surface, most common on hands, but can grow anywhere on the body
- flat wart (*verruca plana*): small smooth flattened wart, can occur in large numbers; most common on the face, neck, hands, wrists and knees

- filiform wart: finger-like wart, most common on the face, especially near the eyelids and lips
- plantar wart (*verruca pedis*): forms a hard, often painful, lump with multiple central black specks; usually on pressure points on the soles of the feet
- mosaic wart: a group of tightly clustered plantar-type warts, usually located on hands or soles of the feet
- genital (venereal) wart (*condyloma acuminatum, verruca acuminata*): occurs on the genitalia
- periungual wart: a cauliflower-like cluster of warts that occurs around the anus.

A wide range of local treatments is available for treating warts. In a systematic revue, Gibbs *et al*. (2002) concluded that no one treatment is strikingly effective and there is little good evidence on the efficacy of these treatments. Such evidence as exists suggests that there may be a beneficial effect from the use of topical salicylic acid and from contact immunotherapy with dinitrochlorobenzene. There is no good evidence to substantiate the efficacy of cryotherapy, although this treatment modality is often used. Likewise, the benefits and risks of 5-fluorouracil, bleomycin, interferons and photodynamic therapy remain to be determined.

Keloid and hypertrophic scars

A keloid is an abnormality of scarring encountered in certain individuals following trauma, including surgical skin incision. It is due to an overgrowth of granulation tissue (collagen type III) which is gradually replaced by collagen type I as the lesion matures. Keloids form firm, rubbery nodular lesions, which characteristically grow beyond the extent of the wound as distinct from the less severe but related condition of hypertrophic scar, which does not cause symptoms other than impaired cosmesis. Keloids are relatively uncommon in white-skinned individuals but are very common in the dark-skinned races, with a reported ratio of 12:1 between dark- and white-skinned individuals.

The distinction between hypertrophic scar and keloid on histology may be difficult because thickened hyalinized collagen, the hallmark of keloid, is not always detectable and -smooth muscle actin (-SMA) is expressed in both conditions (70% in hypertrophic scar and 45% in keloid). The features more commonly encountered in keloids, as reported in a comparative histopathological study (Lee *et al*. 2004), are:

- non-flattened epidermis
- non-fibrotic papillary dermis
- presence of hyalinized keloid collagen (detected in 55% of keloid specimens)
- prominent disarray of fibrous fascicles/nodules
- tongue-like advancing edge underneath normal-appearing epidermis and papillary dermis
- horizontal cellular fibrous band in the upper reticular dermis
- prominent fascia-like fibrous band.

The treatment of keloids is extremely difficult and often problematic. Surgical excision is best avoided as it merely results in further keloid scarring, although scalpel intralesional excision may be needed if non-surgical management has failed. In these cases, excision should be combined with intraoperative or postoperative triamcinolone as recurrence is inevitable with excision alone.

The conservative treatment includes intralesional corticosteroids, which is generally considered as first-line therapy in view of published evidence that the majority of patients obtain substantial improvement with flattening of keloids, although the recurrence rate is high (up to 50%) at 5 years. Silicone gel sheeting is used to treat symptoms (pain and itching) in patients with established keloids. It is also beneficial in the management of patients with evolving keloids. In some of the reported studies treatment with silicone gel sheeting appeared to improve elasticity of established abnormal scars. Cryosurgery is useful in combination with other treatments but it induces hypopigmentation, which limits its use in dark-skinned patients. Radiotherapy has been reported to be beneficial in reducing keloid recurrence after surgical excision. There is, however, concern regarding the potential long-term risks associated with this treatment and for this reason radiotherapy is used only when other therapies have failed. Interferon- injections have been used to reduce recurrence rates postoperatively although there is no firm evidence of their efficacy. Finally, pulsed dye laser treatment appears to induce keloid regression through suppression of keloid fibroblast proliferation and induction of apoptosis. Combination treatment with pulsed dye laser plus intralesional therapy with corticosteroids and/or fluorouracil may be superior to either approach alone.

Malignant and premalignant conditions of the skin

Cancer of the skin as an overall entity is by far the commonest tumour to affect humans. The three main types are SCC, basal cell carcinoma (BCC) and malignant melanoma. These will be dealt with in turn but, initially, the general aetiology and pathology of skin cancer will be considered followed by a description of the premalignant lesions.

Aetiology of skin cancer

There are two major aetiological factors, the HPVs and solar radiation. For some reason, it is usually only the skin of the perineum that is affected by malignant change unequivocally associated with HPV (see Anal intraepithelial neoplasia). However, papillomavirus particles have been found in SCC in other parts of the body.

Both premalignant and malignant lesions of the skin occur more frequently in those areas that are exposed to sunlight. The causative agent is ultraviolet (UV) light, which is classified in accordance with its wavelength (Table 16.1).

UVA (wavelength 320–420 nm) is probably the least damaging to skin but may have an additive effect to that of UVB (wavelength 290–320 nm). UVB is more damaging, and UVC (200–290 nm) is by far the most damaging. Fortunately, relatively little UVC reaches the Earth's surface as it is filtered by the ozone layer, but it can be produced by arc welders and sterilizing lamps. It should be noted that, although sunbeds

Table 16.1 Classification of ultraviolet (UV) light

Class	Wavelength (nm)	Source	Effect
UVA	320–400	Sunlight	Darkening of melanin; phototoxicity
UVB	290–320	Sunlight	Sunburn; carcinogenesis; melanin production
UVC	<290	Arc welders	Carcinogenesis; sterilizing lamps

generally emit UVA radiation, they are still potentially dangerous.

UV light causes dimerization of the adjacent pyrimidines in the epidermal DNA. This decreases the efficiency of the cellular repair mechanisms and thereby may lead to malignant transformation. Malignant and premalignant skin lesions are rare in black populations and common in Australia, New Zealand and South Africa, where the skin of the white migrant North Europeans is exposed to high levels of sunlight over prolonged periods. The absorptive effect of the atmospheric ozone layer is least at the Equator but increases with increasing longitude as the Sun's rays strike the ozone layer more obliquely and have a longer passage through it.

Pathology of solar damage

Solar radiation creates changes in the skin known as actinic change, which involves degeneration of collagen in the dermis and an increase in the amount of elastin. As a consequence, there is failure of the supportive dermal structure and the skin becomes wrinkled. In the epidermal layer actinic change affects all the cells. The epidermis becomes thin and atrophic, melanocytes become irregularly distributed and atypical and there is an increase in the number of suppressor T-lymphocytes.

The epidermal keratinocytes undergo hyperkeratinization and parakeratinization. In the early stages of the dysplastic change, excessive amounts of keratin are produced and the keratin fails to separate from the skin resulting in hyperkeratosis. Clinically, this forms scales. With further accumulation of damage, the pattern changes to the more dysplastic form, parakeratosis. Here the cells fail to mature and, as they reach the surface, they retain their nuclei. In addition, cellular atypia of the individual cells is seen with pleomorphism, hyperchromatism, irregular nuclei and an increase in the ratio of the nuclear to the cytoplasmic areas. Extreme dysplastic changes are described as carcinoma *in situ* and, when the basement membrane is breached, the lesion becomes an invasive cancer.

Actinic keratosis

Actinic keratosis is also known as senile keratosis or solar keratosis. It is most commonly seen in white-skinned individuals and presents clinically as a red scaly patch on the exposed skin of the elderly. The clinical course is slow with regression of some of the lesions and progression of others to malignancy. Histologically, the skin exhibits a variable degree of solar damage as described above. Although the most effective treatment for actinic keratosis is surgical excision, this may prove difficult in large and awkward areas. It may therefore be appropriate to treat the condition conservatively and only advise excision when there are suspicious clinical indications of malignancy, i.e. the onset of ulceration or bleeding.

Carcinoma *in situ*

Carcinoma *in situ* of the skin is sometimes known as *Bowen disease* and it may be difficult to distinguish it from actinic keratosis, as indeed it represents the more severe end of the spectrum of this condition. Histologically, the picture is identical to that of squamous cell carcinoma without penetration of the basement membrane. Clinically, it presents as a roughened reddened patch of skin, which persists over many months or years. Owing to the risk of invasive malignant change, it should be excised, unless there is a compelling reason to treat it expectantly for reasons of size or anatomical position.

Keratoacanthoma

Keratoacanthoma (KA), formerly regarded as a benign self-limiting lesion, is now considered as 'squamous cell carcinoma–keratoacanthoma' (SCC-KA) to reflect the difficulty which often arises in the histological differentiation, as well as the rare but potentially aggressive forms of the disease, especially when multiple. More usually, however, KA behaves as a relatively common low-grade malignant lesion, with an incidence of 104/100 000. It is more common in males (male to female ratio, 2:1), affects elderly patients over 60 years and is rare in dark-skinned individuals. KA originates in the pilosebaceous glands. It may also be a manifestation of the rare *Muir–Torre* syndrome, an autosomal dominant familial cancer syndrome. As the epidemiology is virtually identical to that of invasive SCC and Bowen disease, prolonged exposure to sunlight and chemical carcinogens (industrial workers exposed to pitch and tar) have been implicated, in addition to HPV and an immunocompromised state.

The clinical presentation of KA is that of a rapidly growing skin-coloured or reddish nodule with a central crater containing a keratin plug over a few weeks to months, followed by a slow spontaneous resolution over 4–6 months leaving a scar (Figure 16.6). The majority of KAs occur on sun-exposed areas, e.g. the face, neck and dorsum of the hands and forearms, and are usually solitary. Occasionally KA presents as multiple tumours characterized by rapid local enlargement, and may even metastasize to regional lymph nodes.

Treatment of KA is primarily surgical, with medical treatment being reserved for when surgical intervention is not feasible (multiple scattered lesions, poor surgical risk, etc.). KA should be excised with an adequate margin (3–5 mm) and subjected to detailed histology to exclude invasive SCC.

Medical (conservative) treatment is far from standardized and includes systemic retinoids (isotretinoin) and intralesional cytotoxic agents (methotrexate, 5-fluorouracil, bleomycin). Although successful in the majority, repeat intralesional injections are often needed. KAs are radiosensitive and respond well to low doses of radiation (<10 Gy). Radiation therapy is

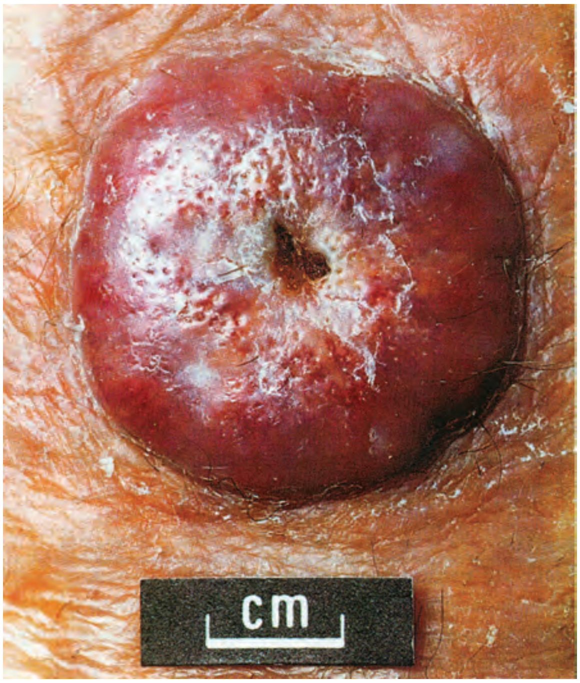


Figure 16.6 Keratoacanthoma. (Courtesy of Dr Sally Ibbotson, University of Dundee, UK.)

used in selected patients with large tumours in whom resection will result in cosmetic deformity and for tumours that recur after excision. Radiation therapy is also used for patients who are unfit for surgery and general anaesthesia. Laser and cryosurgical ablation are used in some centres for small KAs and for lesions which would pose problems with surgical excision by virtue of their location.

Squamous cell carcinoma of the skin

As with carcinoma *in situ*, SCC of the skin tends to occur in exposed areas of white-skinned individuals over the age of 50 years, particularly those who have spent considerable time outdoors. Fair-skinned individuals with light-coloured skin, blue/green eyes and blond or red hair are particularly susceptible.

Aetiology

Aside from prolonged exposure to sunlight, other less important aetiological factors include an inherited predisposition to skin cancer, smoking (SCC of the lip), thermal burn scars, long-standing leg ulcers (Marjolin's ulcer) and immunosuppression by drugs (ciclosporin, azathioprine in transplant recipients). Surveillance is nowadays recommended in all patients receiving long-term medication with immunosuppressive drugs. Infection with HPV causes mucosal SCCs and cuniculatum verrucous carcinoma of the skin of the face. However, HPV very rarely causes common cutaneous SCC. The actual risk of SCC following use of 'biological therapies' for inflammatory and haematological disease in recent years is not known, although there have been documented cases of rapid onset or reactivation of SCC in patients with a past history of the disease, especially in those with risk factors for recurrence. There is good evidence of the efficacy of protective clothing and effective sunblocks in the prevention of both actinic keratosis and SCC.

Pathology

SCC arises from the epidermal keratinocytes. Histologically, abnormal squamous cells are seen growing downwards from the epidermis into the dermis and subcutaneous tissue in strands or sheets. The tumour may be classified as well differentiated, moderately differentiated or poorly differentiated. Typically, the well-differentiated type will exhibit keratinization and formation of keratin pearls composed of concentric layers of cells with increasing keratinization towards the centre. The cells are large and polygonal with eosinophilic cytoplasm, and mitotic figures with few nucleoli. In contrast, the poorly differentiated tumours are made up of anaplastic cells with multiple mitosis and nucleoli and no keratinization. There is a four-part classification of SCC based on differentiation described by Broders. This describes grades 1, 2 and 3, which denote ratios of differentiated to undifferentiated cells of 3:1, 1:1 and 1:3, respectively. Grade 4 indicates a completely anaplastic tumour with no tendency to differentiation.

SCC spreads locally, via lymphatics and via the bloodstream. Local invasion is both horizontal and vertical and, in neglected tumours, can give rise to widespread destruction of the skin and underlying structures. Erosion of large blood vessels can be a terminal event. Lymphatic spread is to the regional lymph nodes in the first instance, and blood-borne spread is to the lungs, bone and brain. However, the ability of SCC for distant metastases is low and only about 3% of tumours will have distant metastases at the time of diagnosis. For some reason, SCCs that arise in chronic scars tend to behave more aggressively, but it is not clear whether this is a biological feature of the disease or whether tumours arising in scars tend to be neglected and present late.

Clinical features

Clinically there are two patterns of growth that are recognized, although these can overlap to a certain extent. The infiltrative type begins as a warty nodule that breaks down to form an ulcer with a rolled edge overlying the surrounding indurated skin (Figure 16.7). The protuberant type produces a raised cauliflower-like lesion (Figure 16.8). The degree of induration felt clinically gives an indication of the depth of invasion of the tumour, as does the extent of ulceration.

Diagnosis

The diagnosis tends to be evident clinically but must be confirmed by either incisional or, whenever possible, excisional biopsy.

Prognosis

The vast majority of SCCs (95%) can be cured by adequate surgical excision, although new tumours may develop subsequently and patients should be informed of this and followed up. Otherwise the best predictor of tumour behaviour in terms of metastatic disease is the vertical tumour thickness (Clark levels), but the surface area of the tumour also has implications for local recurrence. Tumours arising in apparently

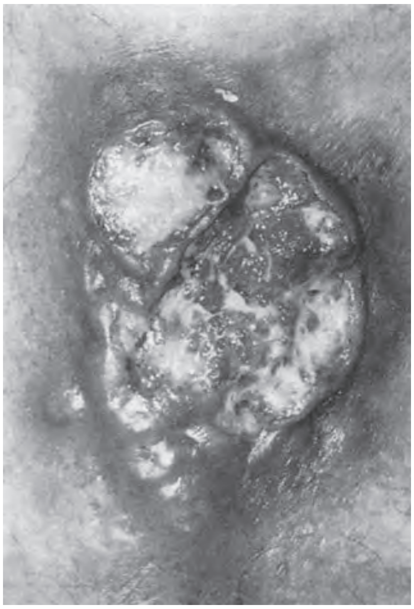


Figure 16.7 Infiltrative squamous cell carcinoma.

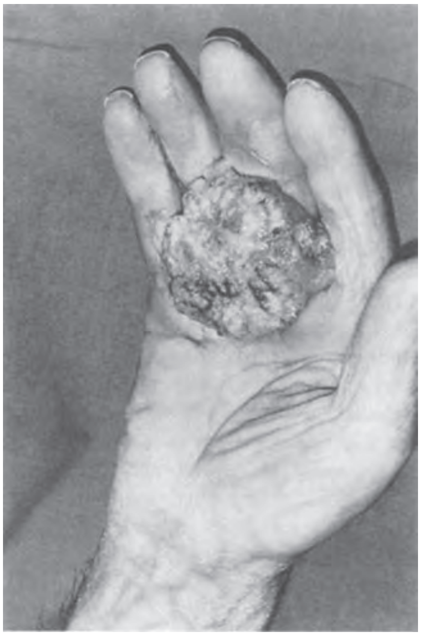


Figure 16.8 Protuberant squamous cell carcinoma.

normal skin have a poorer prognosis in terms of recurrence and distant spread than those arising in skin showing solar damage. Histological differentiation (Broders classification) is also a factor in prognosis as the poorly differentiated tumours have an increased tendency to metastasize. The anatomical site of the tumour is another prognostic factor as tumours of the trunk and limbs have a worse prognosis than those of the head and neck.

Factors which influence the metastatic potential of SCC are:

- *Anatomical site*: greatest potential in SCCs arising in areas of radiation or thermal injury, chronic sinuses/ulcers followed in *descending order* by (1) SCCs in non-sun-exposed sites (perineum, sacrum, sole of the foot), (2) SCC of the ear, (3) SCC of the lip, (4) SCC at sun-exposed areas excluding the lip and ear.
- *Size of tumour at presentation*: tumours >2.0 cm in diameter are twice as likely to recur locally and three times as likely (30%) to metastasize as smaller tumours.
- *Tumour thickness/level of invasion*: tumours >4 mm in depth or extending beyond the subcutaneous tissue (Clark level V) are more likely to recur and metastasize, whereas tumours <2 mm thick rarely metastasize.
- *Histological differentiation and subtype*: poorly differentiated tumours (Broders grades 3 and 4) carry a poorer prognosis with twice the local recurrence rate and three times the metastasis rate than the better differentiated types. Perineural involvement and lymphatic or vascular invasion are associated with an increased risk of recurrence and distant spread. The acantholytic, spindle and desmoplastic subtypes have a poorer prognosis than verrucose lesions.
- *Host immunosuppression*: by drugs or disease.

The Clark levels of invasion are the same as those used for malignant melanoma and are described below.

Treatment

Surgical excision

Surgical treatment remains the treatment of choice for the majority of cutaneous SCCs. Aside from proven efficacy, it allows histological examination of the extent of differentiation of the tumour and objective evidence that the excision has been complete (tumour-free margins). The underlying important consideration which determines the clinical outcome is completed removal of the tumour including the 'in transit' metastases in the environs of large lesions. As previously mentioned, prior diagnosis should be made by excisional biopsy whenever possible and this may serve as treatment if the tumour is completely excised with clear margins. Small lesions may be excised under local anaesthetic, in which case the incision should be planned and marked prior to infiltration with local anaesthetic to avoid mistakes created by the inevitable distortion of the skin.

For low risk SCCs <2.0 cm diameter, surgical excision with a minimum of 4 mm circumferential margin is appropriate and will result in complete removal of the primary in 95% of cases. Tumours >2.0 cm, poorly or undifferentiated tumours, tumours extending into the subcutaneous tissue, and tumours of the ear, lip, scalp or eyelids require a wider margin (minimum of 6 mm). When removing the lesion, it is important to remember that deep clearance is as important as lateral/circumferential clearance and, for this reason, it is important to palpate and inspect the deep margin during the excision. In the limbs, the possibility of extension along fascial planes must be considered and possible perineural spread also taken into account. A history of pain or unusual sensations associated with the tumour may suggest perineural spread.

Irrespective of site, it is essential that a margin of normal tissue is excised on all aspects of the tumour, and, if the tumour is adherent to tendon or bone, this tissue must also be excised. Problems of reconstruction must not compromise the complete clearance of the tumour, otherwise surgical excision will fail. After excision, it is important to mark the specimen in order to enable the histopathologist to orientate the specimen. With this in mind, the specimen should always be accompanied by a diagram showing its site and orientation.

Radiotherapy

Although surgery is the treatment of choice for the majority of SCCs, these tumours are radiosensitive and radiotherapy has an important role in its management. In practice radiotherapy is reserved for tumours which are inoperable by virtue of involvement of vital structures and for postoperative treatment of tumours that have been inadequately excised because of anatomical considerations. Having said this, radiotherapy may be used as the primary treatment modality for all SCCs with the following reservations:

- pretreatment biopsy is essential
- radiation of the exposed skin in young people may increase the risk of future carcinoma and is thus usually contraindicated in this age group
- areas of chronic friction or poor vascularity, e.g. the shoulders, back of the hand, abdominal wall and pretibial region, should not be irradiated as this will lead to chronic skin damage
- treatment of skin around the eye may lead to radiation keratitis.

Because of its superior cosmetic results, radiotherapy is often used instead of surgical excision in SCCs of the lower eyelid, the inner canthus of the eye, the lip and tip of the nose with reported cure rates of >90%.

Other treatment modalities

These include Mohs micrographic surgery, cryotherapy, photodynamic therapy, curettage and cautery and topical therapy. Mohs micrographic surgery consists of microscopically controlled precise excision of the tumour in increasingly deeper layers until the microscopy shows no tumour cells. Nowadays, the fresh tissue modification (without application of zinc chloride paste to fix the tumour *in situ*) is used with excellent results in terms of low recurrence and metastatic rates, such that it is recommended as the treatment of choice for high-risk SCCs, particularly at difficult sites where wide surgical margins may not be possible without inducing deformity of functional impairment. Curettage and cautery is reserved for very small (<1.0 cm), well-differentiated, primary, slow-growing tumours on sun-exposed sites. The lesion is first curetted (for debulking) and then cautery (electrodessication) applied to the wound. The problem with this treatment is that curettage provides poorly orientated material for histological examination and assessment of the completeness of removal is not possible. Good short-tem cure rates have been reported with cryotherapy for small histologically confirmed SCCs. Cryotherapy is contraindicated for recurrent and high-risk tumours. There is no level 1 evidence that other treatments such as photodynamic therapy and topical treatment with imiquimod, 5-fluorouracil paste and interferon- are effective in the long term.

Recurrent disease

If histological examination of the excised specimen reveals involvement of the resection margins, the risk of recurrence exceeds 50%. A second procedure with wider and deeper excision should therefore be undertaken immediately unless this is contraindicated by involvement of vital structures. In this case, radiotherapy is indicated. If recurrence occurs after a lesion has been excised with adequate (clear) histological margins, this indicates deep infiltration along tissue planes. This tends to occur particularly in lesions of the ear and requires a wide excision or radiotherapy. When wide excision is performed, intraoperative frozen section will help to define the extent of excision.

Treatment of metastatic disease

When metastatic deposits in the regional lymph nodes are detected, *en bloc* resection of the lymph node group is indicated and postoperative radiotherapy should be considered for poorly differentiated tumours. Currently there is no evidence for the use of prophylactic block dissection of regional lymph nodes in high-risk large tumours. Some recommend sentinel node biopsy in these cases but again the benefit of this approach remains uncertain.

Metastatic lesions of the bone and brain may be treated by radiotherapy to reduce local symptoms but widespread disease will require chemotherapy. Randomized studies have established the combination of cisplatin and infusional 5-fluorouracil and paclitaxel is effective in palliation (pain relief and improved quality of life) but does not materially improve survival. Phase II studies of a platinum/taxane combination with a third drug have reported response rates exceeding 50%, including 15% complete responses. A number of non-platinum agents are also active including gefitinib and the newer antifolates which target or restore deficient p53. Overall, the 5-year survival in patients with metastatic disease at presentation is 25%. The 10-year survival is 13%, and the 15-year survival is 8%.

Basal cell carcinoma (rodent ulcer)

BCC is the commonest skin malignancy in white-skinned individuals (60% of cutaneous tumours) and is the most common cancer in Europe, Australasia and the USA. The age-standardized incidence of BCC varies considerably from country to country, e.g. 114 per 100 000 population in Wales, 146 per 100 000 in Minnesota, USA, 726 per 100 000 in Australia, etc. The incidence is increasing each year, such that the estimated lifetime risk of BCC in the white populations of North America is 33–39% and 23–28% for males and females, respectively. It is virtually unknown in dark-skinned individuals. BCC arises from epithelial cells and is thought that the cells of origin are from the follicular basal lining of the hair follicle rather than the cells from the basal layer of the epidermis.

Aetiology

The most significant aetiological factors are genetic predisposition and exposure to UV radiation. Increasing age, male sex, skin type I (skin that always burns and never tans), blond hair with blue/green eyes, freckles and sunburn in childhood and adolescence (as opposed to adults), immune suppression, ionizing radiation, burns and arsenic exposure are other recognized risk factors. Multiple BCCs are a feature of the rare *basal cell naevus (Gorlin–Goltz) syndrome*, an autosomal dominant familial disorder characterized by pitting of the palms and soles, jaw cysts, spine/rib anomalies, calcification of the falx cerebri, cataracts and the development of multiple BCCs. Other conditions associated with an increased risk of BCC include albinism and xeroderma pigmentosum. Susceptibility to the development of BCC is determined by a complex interaction between duration, intensity of exposure to UV radiation and certain polymorphic genes.

Pathology

BCCs behave as slow-growing, locally invasive malignant skin tumours which infiltrate local tissues but rarely if ever metastasize. However, untreated they cause local destruction of tissues. BCCs occur on areas of chronic sun exposure, e.g. face, head and neck. Histologically, the tumour consists of clumps of basophilic cells surrounded by a connective tissue stroma. At the periphery of the tumour, columnar cells form a palisade which is the hallmark of the tumour. Mucin is found in the stroma rather than elastin in contradistinction to SCCs. The ratio of cells to stroma varies according to the five recognized types of BCC: noduloulcerative, superficial spreading, morphoeic, pigmented and fibroepithelioma of Pinkus.

Noduloulcerative basal cell carcinoma (cystic basal cell carcinoma, rodent ulcer)

This forms a well-circumscribed tumour with islands of basal cells embedded in a fibroblastic stroma. Centrally the cells are arranged randomly, but peripherally the cells form palisades. The basal cells may show differentiation towards the adnexal structures of the skin or may demonstrate squamous metaplasia.

Superficial spreading basal cell carcinoma

In this variant, there are nests of basiloid cells with peripheral palisading at the level of the basal layer of the epidermis with atrophy of the epidermis above this.

Morphoeic basal cell carcinoma

This type, which accounts for 5% of cases, is the most aggressive. Histologically, it is characterized by the presence of a dense fibrous dermal stroma with scanty strands and groups of cells embedded within it. The tumour cell strands may stream from the under surface of the tumour into the deeper tissues without definite margins. Proliferation of the stroma may also occur in the horizontal plane beneath apparently normal epidermis. This makes for difficulty in judging the margins of the tumour with the naked eye and is responsible for the high incidence of recurrence of this type of BCC. Some of these tumours may be very large at presentation and cause significant disfigurement requiring lengthy plastic surgical reconstruction.

Pigmented basal cell carcinoma

This is similar in histological appearance to the noduloulcerative type but contains melanocytes.

Fibroepithelioma of Pinkus

In this variant, strands of basiloid cells extend into the deeper fibrous stroma but there is no palisading.

Clinical features

Sporadic BCC is rarely encountered before the age of 20 years but increases thereafter with age. All types of BCC develop in the sun-exposed areas with 80% occurring on the head and neck, with the rest mainly on the trunk and lower limbs. The small early lesions are either translucent or of pearly colour and have raised areas with telangiectasia. Otherwise, the clinical features vary with the type. Patients treated for BCCs have an increased risk of developing further (new) lesions. This risk, variously reported at 33–77%, depends on the number of lesions present and location of tumour in the trunk. Additionally, patients with BCCs are at an increased risk of developing malignant melanoma, presumably related to the exposure to UV radiation.

Noduloulcerative basal cell carcinoma

This is by far the commonest form, accounting for approximately 50% of all BCCs. Clinically, it presents as a slowly growing nodule with waxy or pearly appearance associated with telangiectasia. Later in the course of the disease, the skin ulcerates to form a typical rodent ulcer (Figure 16.9).

Superficial spreading basal cell carcinoma

This is also described as superficial cicatrizing or 'field fire' BCC. It presents as a scaling erythematous lesion, usually on the trunk. It does not become indurated, nor does it ulcerate. Instead, it spreads radially while appearing to heal in the centre. The central scar area contains sweat glands but the hair follicles are destroyed producing characteristic alopecia. The tumour is frequently multifocal and is commonly advanced at presentation owing to the insidious nature of its development.

Morphoeic basal cell carcinoma

This presents as a dense greyish yellow plaque that increases in size gradually. Around the eyelids, contraction of the fibrous stroma may result in distortion. The lesion may ulcerate.

Pigmented basal cell carcinoma

Clinically, this type of BCC is similar to the noduloulcerative variety except for a pearly black sheen.

Fibroepithelioma of Pinkus

This presents as a papilloma without scaling or ulceration.

It is important to distinguish between high- and low-risk BCCs, based on their likelihood of recurrence after treatment. Factors indicating higher risk include histological type (morphoeic), size at presentation (>2.0 cm), site (central face), perineural and deep invasion below the dermis, immune suppression and genetic disorders (Gorlin syndrome).

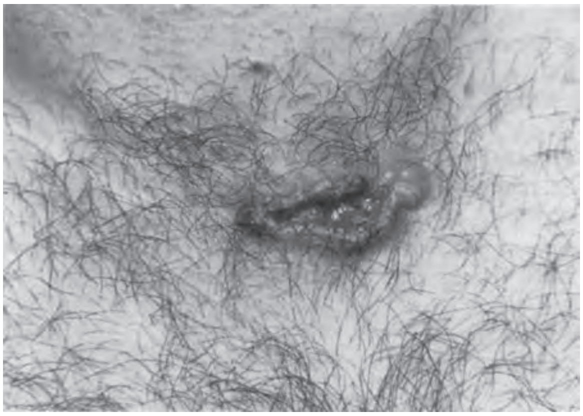


Figure 16.9 Typical rodent ulcer (noduloulcerative basal cell carcinoma).

Diagnosis

The diagnosis of BCC is clinical in the first instance but it must always be confirmed by histology; whenever possible this should be by excision biopsy.

Treatment

The treatment options for BCC are:

- surgical: usually excision with primary closure or with flaps or grafts for large lesions precluding primary closure
- Mohs micrographic surgery
- curettage and cautery
- radiotherapy
- cryotherapy
- photodynamic therapy
- topical fluorouracil
- topical imiquimod.

The choice of treatment in the individual case depends on the size and site of the tumour, its histological nature, the condition of the patient and, in recurrent tumours, the methods previously used.

Surgical excision

This is the standard treatment for the majority of BCCs. A simple excision is usually carried out as an ellipse around the tumour with a margin of 4 mm for noduloulcerative lesions. The excision is planned along Langer's lines to ensure healing with minimal scarring/deformity. In morphoeic and recurrent lesions, the clearance should be significantly larger at 1.0 cm. As with SCC, the excised specimen should be marked and be accompanied by an appropriate diagram to allow the pathologist to orientate the specimen. After excision in most cases, primary closure is possible but large lesions (primary or recurrent) often require partial or full thickness skin grafting. On the face, large lesions may require flap reconstruction by a plastic surgeon after excision. The 5-year cure rate following appropriate surgical excision is 95% or higher.

Mohs microangiographic surgery

When expertise is available for this technique, it is ideal for the treatment of tumours of the central face, morphoeic lesions and recurrent BCCs as it results in very high cure rates (99% for primary lesions and 95% for recurrent BCCs), while providing maximal preservation of normal tissues.

Curettage and cautery

Noduloulcerative BCCs can be treated by curettage and cautery with the wound being allowed to heal by secondary intention. The rationale behind this method is that the soft tumour tissue is scraped out while the firmer more resistant normal skin is left behind. Although curettage and cautery gives cure rates which approximate those of surgical excision, the wound takes much longer to heal and there is no tissue available for histological examination. Curettage is definitely contraindicated for lesions >1.0 cm in diameter, for morphoeic lesions, recurrent tumours and tumours at high-risk sites (central face) as the risk of residual tumour is high.

Cryotherapy

Cryotherapy using a liquid nitrogen probe at minus 50°C for between 30 and 120 seconds effectively destroys BCCs with equivalent cure rates to surgical excision, but, as with curettage and cautery, the wound is left to granulate and heal by secondary intention and no tissue is available for histological examination to confirm complete ablation. Cryotherapy is used selectively, usually for small multiple BCCs.

Topical photodynamic therapy

This is an established technique which has witnessed considerable progress in recent years with the advent of technology based on organic light-emitting diodes (OLEDs). Topical photodynamic therapy is most effective for superficial BCCs, and for these lesions it results in high cure rates with excellent cosmesis. It uses a photosensitizer drug, -aminolaevulinic acid, in a 20% emulsion applied to the tumour, which then takes it up and thereby becomes photosensitized. As a result, it is ablated when exposed to light in the wavelength range of 620–670 nm. The recent OLED-based system has revolutionized photodynamic therapy as, following the application of the sensitizer, a plaster covering containing OLEDs provides the appropriate light irradiation at home (Figure 16.10). The reported average cure rates with photodynamic therapy for BCCs is 87%, but this is lower for nodular lesions (53%), presumably because of the thickness of these tumours adversely affecting the penetration of the light into the deeper layers of the tumour.

Radiotherapy

The overall cure rate of BCC using radiotherapy is about 90% which is of the same order as all other methods used. This option may thus seem attractive as operation is avoided. It does, however

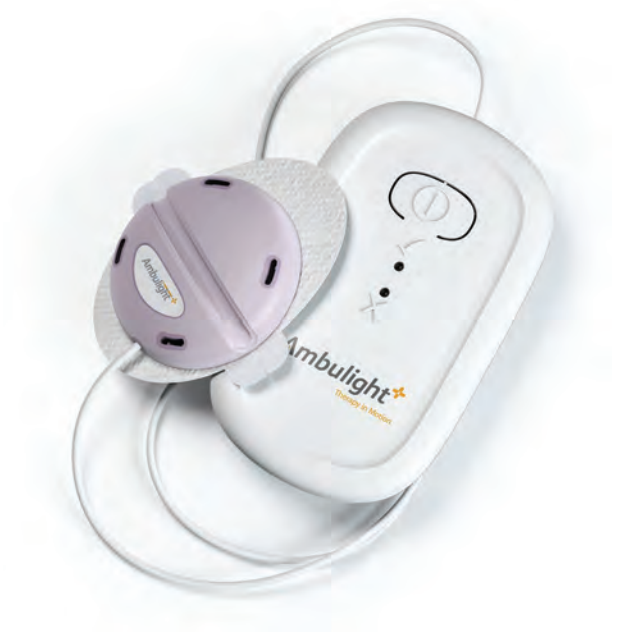


Figure 16.10 Organic light-emitting diode plaster for outpatient photodynamic therapy of superficial basal cell carcinoma. (Courtesy of Professor J Ferguson, Ninewells Hospital and Medical School, Dundee, UK.)

require multiple visits to the hospital as the exposure required is 4000–6000 Gy over several fractions to obtain the best results. It also causes skin atrophy and alopecia. Previous radiotherapy to the area is a contraindication. Additionally, radiotherapy is not recommended for young patients as the late cosmetic results are inferior to those of surgery. In practice, radiotherapy is reserved for elderly patients with extensive lesions in whom major surgery is not appropriate.

Benign pigmented naevi (moles)

Naevi are extremely common and have little clinical significance in themselves. However, they may mimic malignant melanoma or they may be cosmetically unacceptable to the patient. A small percentage of naevi do, however, carry the risk of malignant change.

Aetiology and pathology

Melanocytes, which are responsible for skin pigmentation, arise from the neural crest in the first weeks of embryonic development and then migrate to the skin as well as the meninges, uveal tract and ectodermal mucosa. In the skin, melanocytes are distributed throughout the dermis and epidermis. At birth, most cutaneous melanocytes are present in the basal layer of the epidermis but, over the first three decades of life, some will migrate to the dermis. In the dermis, melanocytes lose their ability to undergo mitosis and therefore have no malignant potential.

Melanocytes may clump together in the basal layer of the epidermis to form a *simple lentigo* or *mole*, which is commonly seen in infants. This may be regarded as the first stage in the progression towards a mature intradermal naevus. As the child grows, the melanocytes aggregate at the dermoepidermal junction. These are called *junctional naevi*. The melanocytes gradually become incorporated into the dermis by fibroblast migration. When this has taken place, naevi consist of a combination of junctional and intradermal aggregates of melanocytes and are termed *compound naevi*. Finally, as the naevus matures, all the melanocytes aggregate in the dermis and an *intradermal naevus* is formed. Thus junctional naevi are common in children and have virtually no malignant potential. In adults, they are rare but do carry a risk of malignant transformation and they are usually found on the palms of the hands and the soles of the feet.

Clinical features

The junctional naevus is either a macule or papule and is brown, smooth and hairless. It may occur anywhere, but when these lesions are seen on the palms or the soles in adults, they may be premalignant. The intradermal type of naevus is a papule which is often hairy and well defined. It may be brown or flesh coloured and is most commonly encountered on the face and the neck.

Diagnosis

The diagnosis is usually self-evident but excision biopsy and histological examination are necessary for definite diagnosis.

Treatment

The only treatment for a naevus is excision but this should only be done when the lesion causes inconvenience to the patient or because of concerns regarding possible malignant change. The latter is suggested by rapid growth, deepening of the pigmentation or bleeding (see below).

Malignant melanoma

Malignant melanoma is the most aggressive of all cutaneous cancers with a marked tendency to spread by both lymphatic and vascular pathways. It predominantly affects white-skinned races, particularly those who are exposed to intense sunlight. The highest incidence in the world is in Queensland, Australia (33 per 100 000). The disease is extremely rare in black-skinned races and, when it does occur, it is found almost exclusively on the sole of the foot, the subungual region or on the mucous membranes. Malignant melanoma is very rare in children and is more common in women than in men in a ratio of 1.5:1. Although predominantly a cutaneous disease, malignant melanoma may rarely originate in mucous membranes (e.g. colorectum, small bowel and oesophagus), and other sites, e.g. choroid plexus of the eye. Not infrequently in these cases, it presents with distant visceral metastases from an unknown primary.

Aetiology and epidemiology

The aetiological factors in malignant melanoma include sun exposure, racial/genetic predisposition, premalignant states and recurrent trauma. The worldwide very significant increase in the incidence of malignant melanoma has been attributed to an increased exposure to UV radiation with the vast increase in charter holidays in the sun by white-skinned residents of wealthy countries. However, recent studies by a Swedish group has established a good time correlation between the roll out of FM/TV broadcasting networks that occurred many years before the boom in holiday charter sun travel. On the basis of this evidence, the authors argue that the increased incidence and mortality of melanoma cannot be solely explained by increased exposure to UV radiation from the sun, as the continuous disturbance of cell repair mechanisms by body-resonant electromagnetic fields may amplify the carcinogenic effects of UV-induced cell damage (Hallberg and Johansson 2004). However, this remains unproven and the Agency for Research on Cancer has concluded that solar radiation remains the main cause of melanoma. In this respect, a review of all case–control studies has concluded that intermittent unaccustomed exposure is more important than age at sunburn. In addition to sunburn (mainly due to UVB radiation), there has been accumulating evidence implicating UVA (and sun beds) in the pathogenesis of melanoma.

The most important premalignant state is the junctional or compound naevus as it appears that melanocytes that remain in the epidermis have distinct malignant potential, the reason for which remains unclear. A minority of malignant melanomas are associated with congenital giant naevi or the *dysplastic naevus syndrome*. Congenital giant naevi are defined as naevi occupying >1% of the body surface area. They may be very large and sometimes cover the trunk, upper arms and upper thighs (bathing trunk naevus). The reported risk of malignant change is such lesions varies from 2% to 28% according to the reported series. Fifty per cent of melanomas which develop in giant congenital naevi do so in the first 5 years of life. The dysplastic naevus syndrome is a familial condition inherited as autosomal dominant with variable penetrance. It is characterized by multiple pale, non-homogeneous naevi with an irregular edge. Most commonly, they occur on the back and start to develop around puberty. The risk of developing malignant melanoma in these individuals is of the order of several magnitudes higher than that of the normal population.

The established risk factors for malignant melanoma are summarized in Table 16.2.

Familial melanoma other than the dysplastic naevus syndrome accounts for 2% of melanomas associated in some cases with two mutated genes (*CDKN2A* and *CDK4*), although these account for a small percentage of familial melanoma. *CDKN2A* codes for two separate proteins (p16 and p14ARF). Both *CDKN2A* and *CDK4* genes control cell mitotic activity. Although families with mutations in *CDKN2A* have an increased risk of melanoma, this varies widely with the geographic region in which they live, presumably because of other factors involved in the aetiology including extent of exposure to sunlight. Mutations in other genes, e.g. *MC1R*, which regulates pigment production by melanocytes and is associated with freckling and red hair, appears to increase the risk of melanoma in individuals both with and without *CDKN2A* mutation.

Pathology

Mutations of the proto-oncogene *BRAF* occur in 50% of malignant melanomas, with the V600E mutation accounting for >90%. This mutation obviates the need for phosphorylation of the T599 and S602 residues – an essential step for the normal activation of *BRAF*. This mutation thus leads to a continuous stimulation of the MAP kinase cascade and hence cell proliferation and dedifferentiation. However, the role of activating *BRAF* mutations with respect to course and stage of melanoma has not been defined, although a recent study has demonstrated that the *BRAF*^{V600E}-dependent expression of a newly identified potential effector gene, *BAALC* (discovered in malignant melanoma cell lines), may be involved as a downstream effector of *BRAF* mutations during the development of malignant melanoma.

Table 16.2 Established risk factors for malignant melanoma

Risk factor	Comment
Number of common moles >2 mm	Risk increases with number of moles >10
Previous history of melanoma	
Family history of melanoma	
Red or light-coloured hair	
Presence of atypical moles	Have irregular borders/pigmentation and >5 mm
Presence of actinic lentigines	Flat brown skin lesion associated with sun exposure
Giant congenital naevi	≥20 cm in diameter
Skin that does not tan easily	
Female sex	Age-standardized female to male ratio 1.5:1.0
Age	Rare in children, thereafter rises to peak in the elderly

Malignant melanomas are classified into four groups in accordance with their clinical presentation:

1. lentigo maligna
2. superficial spreading
3. nodular
4. acral lentiginous.

Superficial spreading melanoma

This is the commonest type of malignant melanoma and typically presents as a change in a pre-existing naevus on the trunk or limbs of an individual aged between 40 and 50 years (Figure 16.11). It appears as a flat, pigmented lesion with an irregular edge and non-homogeneous pigmentation. Histologically, the tumour cells often excite a localized lymphocytic response that appears to induce regression of parts of the tumour. Patients often complain of increases in growth, change in sensation or colour, crusting, bleeding or inflammation. The duration of these symptoms varies from patient to patient from a few months to several years.

Nodular melanoma

This is the second most common melanoma. Again usually arising in a pre-existing naevus, it has the appearance of a nodule growing from a pigmented area of skin usually on the head, neck or trunk of an individual aged between 50 and 70 years (Figure 16.12). Histologically, the tumour has a strong vertical growth component with relatively little radial growth. It usually has a shorter presentation because of a greater tendency to bleed or ulcerate.

Lentigo maligna melanoma

This is a slow-growing, flat, pigmented lesion, sometimes known as Hutchinson's melanotic freckle. It tends to exist for a

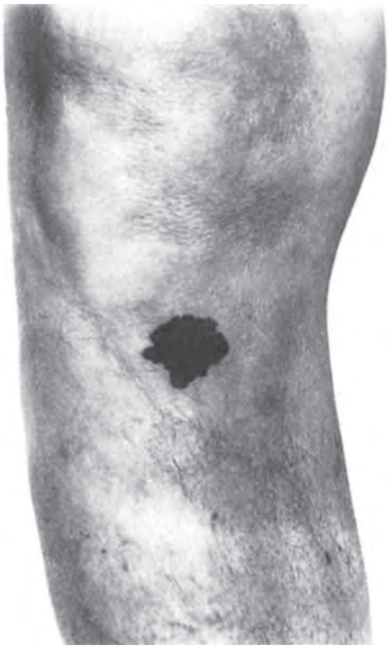


Figure 16.11 Superficial spreading melanoma in the subpatellar region.

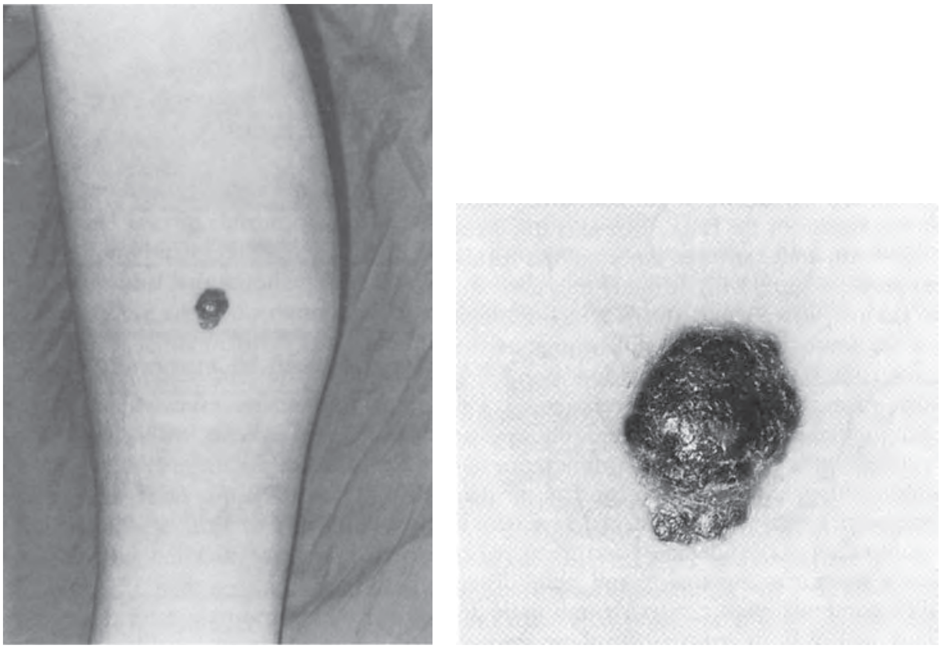


Figure 16.12 (a) Nodular melanoma of the arm; (b) close-up view.

long time in a premalignant state. It is relatively non-aggressive and rarely metastasizes. Lentigo maligna typically develops in sun-damaged skin, usually on the face of elderly individuals.

Acral lentiginous melanoma

This is the least common melanoma. It is used to describe melanomas arising in the nail beds or on the palmar or plantar skin. They tend to grow slowly and seldom metastasize.

Clinical features

The clinical features of a malignant melanoma vary according to the classification previously described in the pathology section. The most important principle to appreciate, however, is that melanoma may develop either in pre-existing naevi or as a *de novo* pigmented lesion in normal skin. The changes in a naevus that suggest malignancy include increase in size, loss of homogeneity with areas of darker or lighter pigmentation, irregularity of outline, nodularity, changes in sensation, crusting, bleeding, ulceration and inflammation. Itching may also be a feature and occasionally the patient may notice satellite lesions.

Diagnosis

Suspicious lesions should be dealt with urgently by excision biopsy which removes the entire lesion with a 2 mm surround of normal skin and underlying subcutaneous fat. Incisional biopsy should be avoided as malignancy may be missed and some clinicians believe that it may encourage metastatic spread, although evidence for this is lacking. Complete excision not only ensures a good histological specimen, enabling a definite diagnosis, but is also essential for the proper staging of the tumour (Clark levels). Another important consideration is the presence/absence of regional lymph node metastases and a careful palpation of the regional nodes is an essential part of the clinical examination.

Prognosis

The prognosis of a malignant melanoma is most accurately assessed by the histological depth of the tumour. In turn, this is measured by the depth of penetration (Breslow thickness) and the Clark levels of skin invasion by the tumour. There is substantial evidence that Breslow thickness is the single most important prognostic variable. However, the Clark level of invasion yields additional prognostic information in patients with a Breslow thickness of <1.5 mm. In particular, Clark level IV in lesions <1.5 mm are associated with significantly higher metastases.

Nodular melanoma is more prone to recurrence than superficial spreading melanomas, although when matched for stage and tumour thickness they have equivalent survival rates. Radial growth phase of a melanoma is a good indication that tumours with this growth characteristic are much less prone to metastases and are associated with a significantly better long-term survival than tumours with vertical growth phase. Ulceration is a well-established bad prognostic factor and is associated with an increased risk of visceral and bony metastases. Additionally, there is an established correlation between extent of ulceration and an increasingly unfavourable prognosis. The association between increased survival and the presence of tumour-infiltrating lymphocytes within the vertical growth phase component remains unproven. Histological evidence of lymphatic invasion is associated with in-transit metastases and early (within 12 months) locoregional cutaneous recurrence. Likewise, the presence of satellites also increases the risk of both local recurrence and metastatic spread. Various models have been used to predict survival in the individual patient. The most commonly used is the Cochran model, which predicts 3- to 10-year survival and is based on clinical data (age, sex and anatomical site) in combination with the Breslow thickness and the breadth of epidermal ulceration. The same model can be modified to calculate risk of recurrence by inclusion of histogenetic subtype and the regional lymph node status.

Breslow thickness

The Breslow thickness is measured from the granular layer of the epidermis to the deepest melanoma cell of the lesion in a fixed specimen. Breslow separated melanomas into three bands according to the prognostic significance of the depth of penetration. Tumours <0.76 mm in depth were accompanied by a 100% long-term survival; those between 0.76 mm and 1.5 mm have an unfavourable prognosis; and tumours >1.5 mm are uniformly associated with a bad prognosis. Thus prognosis deteriorates with increasing tumour thickness. However, in thin tumours which have stimulated an immune response resulting in the regression of tumour cells (superficial spreading melanoma), measurement of the Breslow thickness may be unreliable. In these cases, survival will be less than predicted for the apparent thickness.

Clark levels

Clark classified five levels of invasion of the skin by malignant melanoma in terms of the histological layers of the skin. The system requires accurate histological assessment of the penetration of the tumour through the defined layers of the dermis. It is subject to some interobserver variability and, for this reason, it is somewhat less accurate than Breslow thickness, which remains the gold standard in determining prognosis in the individual patient. The Clark levels are:

Level I	Tumour cells confined to the epidermis above the basement membrane.
Level II	Tumour cells present in the papillary dermis.
Level III	Tumour cells penetrating through the full depth of the papillary dermis to the level of the reticular dermis.
Level IV	Tumour cells in the reticular dermis.
Level V	Tumour cells in the subcutaneous fat.

Staging

In addition to the Breslow thickness and the Clark level of skin involvement, the staging of melanoma has to take into account metastatic spread (nodal and distant metastases). The simplest staging system is the McNeer and Dasgupta staging:

Stage 1	Primary tumour only with no evidence of metastases.
Stage 2	Primary lesion with metastatic spread to the regional lymph nodes detected on clinical examination.
Stage 3	Primary lesion with metastases to more than one lymph node group or distant metastases.

However, more detailed systems that enable better prognostic information and better management are used nowadays. This includes the American Joint Committee on Cancer (AJCC) and the tumour–node–metastasis (TNM) classification. The AJCC staging is as follows:

Stage 1A	Localized disease <0.76 mm or Clark level II.
Stage 1B	Localized disease 0.76–1.5 mm or Clark level III.
Stage IIA	Localized disease 1.5–4.0 mm or Clark level IV.
Stage IIB	Localized disease ≥4.1 mm or Clarke level V.
Stage III	Melanoma with involvement of regional lymph nodes.
Stage IV	Melanoma with advanced regional node involvement, multiple regional node group involvement or distant metastases.

The best characterization of the stage of the melanoma in the individual patient is, however, provided by the TNM system (Tables 16.3–16.5), which aside from indicating the appropriate management also enables reliable prognosis, certainly in terms of survival rate.

Treatment

The treatment of malignant melanoma is essentially surgical. The initial excision biopsy should be carried out with a sufficient margin to ensure complete local clearance. Following histological confirmation of the diagnosis and measurement of Breslow thickness, further treatment may be planned. The basic principle is wide excision to prevent local recurrence caused by micrometastases present in the skin at the time of the primary excision, but there is still some controversy surrounding the

Table 16.3 Tumour (T) classification

Breslow thickness	Ulceration	Stage	5 year survival
T1 ≤1.0 mm	a: Without ulceration and level II/III	1A	95%
	b: With ulceration or level IV or V	1B	90%
T2 1.01–2.0 mm	a: Without ulceration	IIA	90%
	b: With ulceration	IIB	78%
T3 2.01–4.0 mm	a: Without ulceration	IIA	78%
	b: With ulceration	IIB	65%
T4 >4.0 mm	a: Without ulceration	IIB	65%
	b: With ulceration	IIC	45%

Table 16.4 Node (N) classification

No. of metastatic nodes	Nodal metastatic mass	Stage	5 year survival
N1: 1 node	a: Micrometastases	IIIA	70%
	b: Macrometastases	IIIB	59%
N2: 2 or 3 nodes	a: Micrometastases	IIIA/B	63%
	b: Macrometastases	IIIB/C	59%
N3: 4 or more nodes or matted nodes or a combination of in-transit metastases/ satellites/ ulceration with nodes		IIIC	27%

Table 16.5 Metastasis (M) classification

Site	Lactate dehydrogenase	Stage	5 year survival
M1a: Distant skin, subcutaneous or nodal metastases	Normal	IV	15–20%
M1b: Lung metastases	Normal	IV	5–10%
M1c: All other metastases/ visceral		IV	8–10%

margin required to ensure complete clearance. Studies on comparison of 1 cm and 3 cm margins for tumours up to 2 mm thick have documented no overall difference in survival between the two groups. A 1.0 cm margin is thus considered adequate for melanomas <1 mm thick. For lesions between 1 and 2 mm thick a width excision of 1–2 cm is, however, considered necessary. The AJCC recommends the following:

- For pTis (melanoma *in situ*): a surgical excision margin of 2–5 mm to achieve complete histological excision.
- For pT1 (melanoma 0–1 mm thickness): a surgical excision margin of 1.0 cm.
- For pT2 (melanoma 1–2 mm thickness): a surgical excision margin of 1.0–2.0 cm.
- For pT3 (melanoma 2–4 mm thickness): a surgical excision margin of 2.0 cm.
- For pT4 (melanoma >4 mm thickness): a surgical excision margin of 2.0 cm.

However, a randomized controlled trial (RCT) (Meirion *et al.* 2004) involving 900 patients with melanoma undergoing excision randomly assigned to 1.0 cm or 3.0 cm margins showed that excision with a 1.0 cm margin was associated with significantly increased (168 recurrences) locoregional recurrences compared with the group with 3 cm margins (142 recurrences). In this RCT, there were 128 deaths attributable to melanoma in the group with 1 cm margins, as compared with 105 in the group with 3 cm margins, but this did not reach significance. Thus, a 1.0 cm margin of clearance is inadvisable for high-risk melanomas (defined by a tumour thickness of at least 2.0 mm) as it is associated with a significantly greater risk of regional recurrence than is obtained by a 3.0 cm margin.

Obviously the width of excision at sites of aesthetic and functional importance requires full clinical consideration but must not compromise complete clearance. The extent of the deep excision margin is equally important in all cases and should always include adipose tissue down to, but not including, the deep fascia. Following excision, the majority of wounds can be closed by primary skin closure but wide excisions will require plastic surgical reconstruction.

There is some controversy surrounding the treatment of regional lymph nodes in patients with stage I disease, i.e. those with localized disease with no clinically involved nodes. Reports on elective regional lymph node dissection in such patients have shown that tumour may be present in 5–25%. For this reason, some surgeons will perform an elective regional lymph node clearance in those patients with tumours thicker than 1.5 mm, even when the regional lymph nodes are impalpable. For patients with clinical lymph node involvement but no apparent distant metastases, it is generally agreed that block excision of regional lymph nodes should be performed. Following this, the 5 year survival is related to the original Breslow thickness, with 80% survival in patients with melanomas <3.5 mm and 27% survival in those with tumours with thickness >3.5 mm.

Probe-directed sentinel node dissection/biopsy is now well established in the management of patients with malignant melanoma. In this technique, patients with clinical stage I and II melanomas (i.e. without palpable lymph nodes) undergo intraoperative mapping of the regional lymph nodes. This is done by preoperative intradermal injection in the region of the tumour (or excision area) with ^{99m}Tc-labelled sulphur colloid or human serum albumin. The draining lymph nodes are then detected by a hand-held gamma probe (Figure 16.13) and the 'hot' node(s) identified and excised. If these sentinel nodes are proved by histology to contain tumour deposits, a block dissection of the regional node is carried out. There is good evidence that sentinel lymph node biopsy accurately determines the presence/absence of metastasis within the regional lymph nodes and is a useful staging tool in melanomas >1 mm. In thick melanomas (>4 mm) sentinel node biopsy can identify a subset of good prognosis melanomas which are node negative.

In patients with palpable nodes, fine needle aspiration (FNA) cytology should be used in the first instance to confirm the diagnosis of metastatic disease before block dissection is performed. Biopsy is performed if the FNA is inconclusive. In the groin, the block dissection should include the superficial inguinal nodes and the deeper obturator and iliac nodes (radical ilioinguinal) as this offers a survival advantage over the superficial inguinal dissection, although it is accompanied by increased morbidity and thus should be carried out by experts. The 10 year survival after complete ilioinguinal node dissection depends on the number of nodes involved and varies between 20% and 45%. Head and neck melanomas have the most variable pattern of lymph node metastasis and require a variety of types of neck dissection that may include the parotid or the posterior occipital chain nodes. There is no evidence that adjuvant radiotherapy after radical block dissection improves survival.

Other forms of treatment

These include isolated limb perfusion with melphalan and immunotherapy as adjuvant therapy for patients with stage II and III disease

Isolated limb perfusion

This is a surgical technique that effectively isolates the limb circulation enabling high-dose chemotherapy with melphalan to the limb within minimal systemic leakage and toxicity. It is used for the treatment of major limb recurrences. Isolated

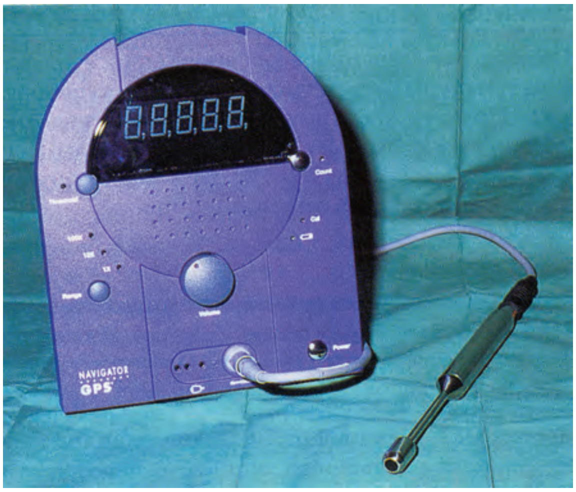


Figure 16.13 Hand-held gamma probe used for sentinel node biopsy.

limb perfusion is a highly specialized treatment that is carried out in regional centres only. A large prospective multicentre RCT showed that prophylactic isolated limb perfusion with melphalan is not indicated in high-risk primary limb melanoma.

Immunotherapy

This is based on the established observation that many primary melanomas undergo partial regression and, occasionally, complete regression. The two approaches proposed in patients with stage II or III disease are non-specific immune stimulation with interferon- and tumour-specific vaccines. Although several studies with interferon- have documented longer disease-free intervals after surgery, no trial has reported a significant increase in overall survival and, especially in view of the toxicity including marrow depression with adjuvant interferon, the current recommendation is that interferon therapy should not be used in patients with stage II or III disease other than in a trial setting.

Melanoma tumour vaccines are still in the research phase. One RCT (the Vaccinia Melanoma Oncolysate Trial) did not show an advantage for the vaccine. Another large RCT compared vaccinia melanoma cell lysates with no immunotherapy and found that vaccinia melanoma cell lysates did not significantly improve overall survival or relapse-free survival. Thus, as with interferon, melanoma vaccines are not in established routine practice for stage II or III disease but recruitment of suitable patients into vaccine trials is recommended.

Management of stage IV disease

This includes surgical excision (metastasectomy), chemotherapy and radiotherapy. Metastasectomy is a good option for patients with distant skin, node or visceral metastases. Surgical excision of single or localized metastases improves survival. However, the proportion of patients suitable for excision is limited to 10–25% of patients, with the 5-year survival varying from 14% to 33% in reported series, depending on the site (skin and lung being favourable), the number of metastases and the disease-free interval. Surgery is also indicated for spinal cord compression, and these patients require urgent referral to a neurosurgical centre as they do not benefit from chemotherapy or radiotherapy.

Dacarbazine is the standard single agent of choice in stage IV melanoma. Multiple drug regimens, including those with tamoxifen and interferon- , do not improve survival compared with single agent dacarbazine, which has an overall response rate of 20%. A meta-analysis comparing dacarbazine with combination chemotherapy, with or without immunotherapy, found no statistically significant benefit for multiple drug regimens compared with dacarbazine alone. Interferon- combined with dacarbazine does increase the response rate to 50% but does not improve survival. In a RCT, the Dartmouth regimen (dacarbazine, cisplatin, carmustine and tamoxifen) was compared with single agent dacarbazine. This showed significantly greater toxicity with the combination regimen though the response rate was higher (18.5% versus 10.2%), but again was not accompanied by a survival advantage. More recently, an oral preparation of dacarbazine (temozolomide) has been introduced and shown to have equivalent efficacy and better central nervous system penetration.

Radiotherapy is used for palliation of pain from bony deposits. Single-dose radiotherapy of at least 8 Gy provides effective palliation. In patients with cerebral metastases with good performance status and favourable response to corticosteroids, surgical resection is indicated in the absence of systemic disease. In patients with inoperable cerebral metastases, radiotherapy with systemic corticosteroids is used for palliation.

Tumours of fibrous tissue

Fibroma

This is a benign encapsulated fibrous nodule which may either form a soft subcutaneous nodule (fibroma mole) or a pedunculated lesion of the oral mucosa (fibroma durum). Both are treated by excision. A third variety, the pleomorphic fibroma, presents in adult patients as a slow-growing, dome-shaped papule most commonly in the limbs followed by the trunk, and head and neck. Histologically, it forms a well-circumscribed lesion composed of dense collagenous tissue with scant spindle, stellate and pleomorphic multinucleated giant cells with few if any mitotic figures. The stroma may exhibit myxoid change. Again treatment is by local excision.

Benign fibrous histiocytoma

This condition is also known as sclerosing haemangioma or dermatofibroma and forms a benign lesion usually in the arms and legs in adults. The lesions seldom grow beyond 1.0 cm and histologically they are made up of thin-walled capillary vessels surrounded by fibroblasts and histiocytes. Treatment is by local excision.

Malignant fibrous histiocytoma

Malignant fibrous histiocytoma (MFH) is the most common soft-tissue sarcoma occurring in late adult life and accounts for 20–24% of soft-tissue sarcomas. MFH originates either from the deep fascia or skeletal muscle usually in the limbs (70–75%), but may arise in the retroperitoneum where it usually presents late. MFH has also been reported in internal organs, e.g. lung, kidney, bladder, stomach, small intestine, and other sites including the orbit and central nervous system.

Pathology

The histogenesis of MFH is uncertain. The tumour contains both fibroblast and histiocyte-like cells in varying proportions, together with spindle and rounded cells. The several histological subtypes of MFH are:

1. *Storiform/pleomorphic MFH*: undifferentiated high-grade pleomorphic sarcoma in the World Health Organization (WHO) classification (most common subtype).
2. *Myxoid MFH*: myxofibrosarcoma in the WHO classification.
3. *Giant cell MFH*: undifferentiated pleomorphic sarcoma with giant cells in the WHO classification.

4 *Inflammatory MFH*: undifferentiated pleomorphic sarcoma with prominent inflammation in the WHO classification.
5 *Angiomatoid MFH*: tumour of uncertain differentiation (generally affects children and young adults and shows indolent behaviour with low metastatic potential).

Clinical features

The peak incidence of MFH is in the fifth and sixth decades but varies considerably. Although rare in children, the angiomatoid subtype may occur in patients younger than 20 years. MFH is more common in males with an male to female ratio of 2:1. It presents with an enlarging, painless soft-tissue mass most commonly in the muscle compartment of the thigh. Rapid enlargement is seen in pregnant females. Rarely, the patient may exhibit symptoms of episodic hypoglycaemia or has/develops a haematopoietic disorder (Hodgkin and non-Hodgkin lymphomas, multiple myeloma). Retroperitoneal MFH presents with constitutional symptoms, including fever, malaise and weight loss. In these patients, the tumour is usually large at presentation and may cause displacement of the organs: bowel, kidney, ureter and/or bladder. Distant metastases occur most commonly to the lung (90%) and less frequently to bone (8%). MRI is the imaging modality of choice for the diagnosis and staging of MFH because of its ability to provide superior contrast between tumour and muscle. CT is used for patients in whom MRI is not possible for any reason. As no single imaging technique can provide a specific histological diagnosis, biopsy is essential for the definite diagnosis.

The overall 5-year survival of patients with MFH ranges from 36% to 58% but is considerably worse for retroperitoneal tumours (15–20%). Otherwise, prognosis depends on the clinical stage, tumour grade, histological subtype, size of tumour and presence of distant metastases. Whereas patients with low-grade MFH have 10-year survival of 90%, this falls to 20% with high-grade lesions.

Treatment

Surgery is the mainstay of treatment. For MFH of the extremity, the surgical options are twofold: limb-sparing resection and amputation. As these result in equivalent survival rates, currently limb-sparing surgery is overwhelmingly the favoured treatment but is only indicated if it allows complete removal with reasonable residual limb function. With limb preservation, the resection for MFH may be either *wide* (removal of the entire tumour with margin of normal tissue) or *radical* (the entire compartment containing the tumour is removed).

External beam radiotherapy is administered routinely in all high-grade tumours as one RCT confirmed abolition of local recurrence compared with a 20% local recurrence in limb salvage surgery alone. The role of chemotherapy in the treatment of MFH remains uncertain, and currently chemotherapy is reserved for patients with metastatic disease.

Desmoid tumour (aggressive fibromatosis)

The term desmoid is derived from the Greek word *desmos*, which means a band or a tendon, and desmoid tumours arise from fascial sheaths or musculoaponeurotic structures. They are also known as aggressive fibromatosis, which better describes their cellularity and aggressive local behaviour. They originate from myofibroblasts in myoaponeurotic structures throughout the body. Desmoid tumours are infiltrative, usually well-differentiated tumours but have a marked tendency for recurrence. Although most commonly originating from the fascia of the rectus abdominis in women after childbirth and in incisional scars after previous abdominal surgery, desmoid tumours may arise in any skeletal muscle compartment. Desmoid tumours do not usually involve the overlying skin.

Sporadic desmoid tumours are rare and account for 0.03% of all neoplasms. Desmoid tumours are much commoner in patients with familial adenomatous polyposis (FAP) of the colon, the prevalence in these patients being variously reported at 9–13%. This genetic predisposition of patients with FAP to desmoid tumours is independent of the *APC* mutation.

Clinical features

Desmoid tumours are uncommon. The estimated incidence in the general population is two to four per million people per year. They occur in patients aged between 15 and 60 years. Desmoids are commoner in females than in males (female to male ratio of 1.2:1.0). The average age of reported cases is 41 years, with an age range of 10–60 years. The commonest sites of desmoid tumours are the anterior abdominal wall and shoulder girdle. Retroperitoneal tumours are more common in familial polyposis coli. There is usually a history of trauma (surgical intervention is present in 25% of cases). A biopsy is considered necessary for establishing the diagnosis.

From a clinical standpoint, desmoids commonly form smooth, firm, initially painless, mobile tumours arising in operation scars, e.g. after Caesarean section. They do not usually affect the overlying skin and their development should alert the surgeon to the possibility of FAP and hence the need for a detailed family history. Both intra- and extra-abdominal (limbs, scrotum) desmoids are rare. Intra-abdominal desmoid tumours remain asymptomatic until they reach a large size and then present with evidence of compression of adjacent organs: intestinal tract, ureter or vascular tree.

Aetiology

The exact cause of desmoid tumours is not known but trauma, hormonal and genetic factors appear to be involved. The role of hormonal factors is suggested as these tumours most commonly arise in young women during or after pregnancy and may regress with oral contraception, with tamoxifen treatment and during menopause. The estimated risk of developing a desmoid tumour in patients with FAP is between 4% and 20%. The desmoid tumours that occur in patients with FAP frequently develop in areas of previous surgical procedures, often in the scar of a Caesarean section. Trauma related to pregnancy and surgery and exposure to elevated hormone levels may both be contributory.

Treatment

The treatment of desmoid tumours consists of surgical excision, which must ensure tumour-free margins to reduce the marked tendency for recurrence (up to 50%). Medical therapy may be considered as an initial alternative, particularly for tumours that involve the intestines, blood vessels, nerves or organs. Other treatment modalities that are used in selected patients include radiotherapy and chemotherapy.

Medical therapy

Mostly this is with non-steroidal anti-inflammatory drugs, e.g. sulindac usually in combination with an antioestrogen (tamoxifen). Although this usually results in detectable symptomatic improvement, significant shrinkage occurs in only a small minority of patients. Desmoids may respond to the tyrosine kinase inhibitor imatinib, presumably because of expression of one of imatinib's molecular targets – the platelet-derived growth factor receptor. Antifibrotic agents, e.g. pirfenidone, have been used recently and although promising the reported experience is still very limited.

Radiotherapy

This is an effective alternative option to surgery for some patients and may be used as an adjunct to surgical excision. Radiotherapy is, however, contraindicated in intra-abdominal desmoids in view of their large size at the time of presentation and involvement of vital structures.

Chemotherapy

Chemotherapy may be effective in patients with unresectable tumours. It is used if tumours do not respond to tamoxifen and/or sulindac. Low-dose chemotherapy with methotrexate and vinblastine produces good response rates, particularly in children. Liposomal doxorubicin has shown promising response rates. Low-dose chemotherapy is usually administered for a year and then stopped. In view of systemic toxicity, aggressive high-dose chemotherapy with doxorubicin and dacarbazine is reserved for patients unresponsive to other therapies or with very rapidly growing/symptomatic tumours.

Other options

Other treatment options include intralesional therapy with acetic acid and radiofrequency ablation but these have been limited to a few centres without sufficient reported data to adjudicate their efficacy in shrinking the tumour (see also Chapter 21).

Fibrosarcoma

Fibrosarcoma (Figure 16.14) is a malignant neoplasm of mesenchymal cell origin in which histologically the predominant cells are fibroblasts. It forms 5% of all primary bone sarcomas, is locally invasive and prone to distant metastases. The tumour affects men and women with equal frequency and can occur at any age, but most commonly presents between the third and sixth decades. In older patients, fibrosarcoma may originate from a pre-existing benign lesion, e.g. chronic osteomyelitis, burn scars, in areas of previous irradiation and fibrous dysplasia.

On histology the tumour is classified as low or high grade. Low-grade fibrosarcomas have abundant collagenous stroma, few cells and mitotic figures. These are the better differentiated tumours and have a relatively good prognosis. In contrast, high-grade fibrosarcomas have abnormal fibroblasts arranged in a herringbone pattern with giant cells and mitotic figures and exhibit a poor prognosis. The prefix A (stage I) is used for tumours that remain within the anatomical compartment at the time of presentation, and B (stage II) when the fibrosarcoma involves adjacent compartments. Stage III is used to describe tumours of any grade that have metastasized. The majority of fibrosarcomas present at stage IIB. Distant spread from fibrosarcoma is to other bones, lymph nodes, brain, subcutaneous tissue, muscle and visceral organs, but rarely to the lung.

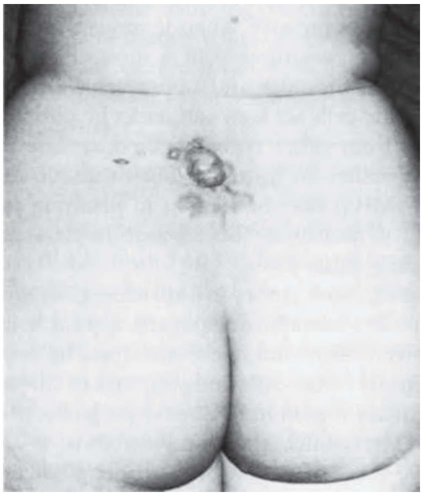


Figure 16.14 Fibrosarcoma.

Surgical resection is the standard treatment for fibrosarcoma. Chemotherapy is administered to patients with stage IIB or III disease usually before surgery (neoadjuvant chemotherapy). Radiotherapy is only used for palliation. Surgery consists of amputation or limb salvage surgery, which is possible with advances in adjuvant chemotherapy, surgical technique and implant development. Amputation is, however, indicated for unresectable lesions, local recurrence, failure of previous reconstruction and for certain tumour locations which preclude functional reconstruction. Although amputation avoids the complications associated with limb salvage (non-union, implant infection and implant failure) and has a lower local recurrence rate, it does not impart a survival advantage over limb salvage surgery.

Prognosis in fibrosarcoma is largely dependent on the histological grade. Survival rates from 50% to 80% have been reported in low-grade and surface fibrosarcomas. Other factors associated with a poor survival include age over 40 years and advanced stage of disease at presentation.

Dermatofibrosarcoma protuberans is a rare form of fibrosarcoma which presents clinically as one or more red or bluish dermal nodules usually on the trunk or shoulders. These tumours tend to occur in young adults and although they grow slowly they may suddenly start to enlarge and ulcerate. They are extremely locally aggressive but metastatic disease is relatively rare. On histology there are non-collagenous fibroblasts arranged in bundles around a central acellular area producing a characteristic cartwheel appearance. Treatment is by wide local excision.

Tumours of muscle

Rhabdomyoma

There are very rare benign lesions of striated muscle and consist of two types:

- true benign tumours:
  - adult
  - fetal
  - genital
- hamartomas:
  - cardiac rhabdomyoma
  - rhabdomyomatous mesenchymal hamartoma of skin.

The adult rhabdomyoma occurs predominantly in the head and neck region including the tongue of men aged 25–40 years as a round or polypoid mass. Fetal rhabdomyoma occurs in children aged between birth and 3 years also in the head and neck. The genital type affects the vagina or vulva of middle-aged women and may cause dyspareunia.

Cardiac rhabdomyoma arises in fetal life and involves both ventricles and interventricular septum. It presents in children and may be associated with tuberous sclerosis of the brain, sebaceous adenomas and hamartomas in other organs. Rhabdomyomatous mesenchymal hamartoma presents in newborns and infants as a lesion in the subcutis, which on histology exhibits skeletal muscle bundles, fibrous tissue, fat and nerves.

Treatment of rhabdomyomas is by local excision.

Rhabdomyosarcoma

Rhabdomyosarcoma is a malignant mesenchymal tumour that arises from primitive muscle cells. It is a disease of children and adolescents, with the vast majority of tumours (90%) occurring below 15 years, the disease being very rare in adults. Several distinct histological types of rhabdomyosarcoma that influence the clinical course and prognosis of the disease are recognized. The cause of rhabdomyosarcoma is unknown, although genetic factors are probably involved as several genetic syndromes are associated with increased prevalence of the disease, e.g. neurofibromatosis, *Li–Fraumeni* syndrome (germline mutation of the tumour suppressor gene *p53*), *Gorlin's basal cell naevus syndrome, Beckwith–Wiedemann syndrome* and *Costello syndrome.*

Pathology and types of rhabdomyosarcoma

Several distinct histological types are recognized and these influence the prognosis. The most common type accounting for 55% of cases is *embryonal rhabdomyosarcoma*, which carries the best prognosis and is frequently found in the head and neck and in the genitourinary organs. *Botryoid* (resembling grapes) *rhabdomyosarcoma* is a subtype of rhabdomyosarcoma that affects hollow organs, e.g. urinary bladder and the vagina in young girls. *Spindle rhabdomyosarcoma* is the usual type affecting the testicular region in young boys. Other types include *alveolar rhabdomyosarcoma*, which has septa and thus resembles pulmonary alveoli, and *undifferentiated sarcoma*. The botryoid variant arises in mucosal cavities, e.g. bladder, vagina, nasopharynx and middle ear, whereas the alveolar type is most commonly encountered in the extremities. The two types associated with a poor prognosis are the alveolar and undifferentiated rhabdomyosarcoma. Distant metastases occur predominantly in the lungs, bone marrow, lymph nodes, breasts and brain.

Clinical features

The overall annual incidence of rhabdomyosarcoma is six cases per 1 000 000. The disease is more common in males, with a male to female ratio of 1.4:1. The most common sites of the disease are the head and neck (28%), followed by limbs (24%) and genitourinary tract (18%). Other less common sites include the trunk (11%), orbit (7%) and the retroperitoneum (6%). Children aged 2–6 years tend to develop rhabdomyosarcoma either in the head or the genitourinary system. In contrast, in adolescents rhabdomyosarcoma affects the limbs, trunk and testicular/scrotal regions.

Rhabdomyosarcoma usually presents as an expanding mass, the location of which determines the symptoms in the individual patient. Thus, orbital tumours cause proptosis and dysconjugate gaze; vaginal rhabdomyosarcoma presents as a protruding polypoid mass; it is a painless testicular/paratesticular mass in young boys; and is an initially painless enlarging mass of the extremities in adolescents of either sex, etc. Metastatic disease presents usually with bone pain or respiratory difficulty/breathlessness from pulmonary metastatic disease and pleural effusion. In advanced disease, anaemia, thrombocytopenia and neutropenia are usually present and, for this reason, disseminated rhabdomyosarcoma may mimic leukaemia.

The clinical staging of rhabdomyosarcoma determines the treatment in the individual case. The current clinical system of staging is based on extent of tumour and on residual disease after resection (Table 16.6).

Approximately 60–70% of newly diagnosed patients with non-metastatic disease can be cured with combined modality therapy. Despite aggressive multimodality treatment, <20% of patients with metastatic rhabdomyosarcoma are currently cured of their disease. Analysis of data from the Surveillance, Epidemiology, and End Results (SEER) programme indicates that mortality is related to age, site and histology, the 5-year survival being highest in children aged 1–4 years (77%) and

Table 16.6 Clinical system of staging of rhabdomyosarcoma

Clinical group	Extent of disease
IA	Tumour confined to the site of origin; resected completely
IB	Tumour extending beyond the site of origin but resected completely
IIA	Tumour with microscopic localized disease after resection
IIB	Extensive tumour + regional node spread, resected completely
IIC	Extensive tumour + positive regional nodes with microscopic residual disease
IIIA	Extensive tumour with positive lymph nodes, gross residual disease after biopsy only
IIIB	Locally extensive residual disease after ≥50% debulking
IV	Any tumour with distant metastases

worst in infants and adolescents. Orbital and genitourinary sites are associated with good 5 year survival (80%), whereas with tumours of the extremities, retroperitoneum and trunk survival is lower and averages 50%.

Treatment

The treatment of rhabdomyosarcoma depends on the stage of the disease, but is always multimodal with surgery, radiotherapy and prolonged chemotherapy.

Surgery

Complete surgical resection is recommended provided it is not mutilating. When complete resection is not possible, neoadjuvant chemotherapy is initiated after the procurement of a biopsy and this is often combined with radiotherapy. Second-look surgery is undertaken in patients who do not achieve a complete remission, with resection of residual disease. However, it is not known whether this improves survival.

Radiotherapy

Radiation therapy is used for sites such as the head and neck or pelvis, where complete surgical excision is not possible. For residual microscopic disease, doses of 4000–4500 cGy are used but a higher dose (5000 cGy) is considered necessary for gross residual disease and for tumours >5 cm.

Chemotherapy

This is necessary as even localized rhabdomyosarcoma usually has micrometastatic disease. The development of adjuvant and neoadjuvant therapy has increased survival in patients with localized disease to approximately 60%. The cytotoxic drugs that are active and used in different combination chemotherapy regimens are vincristine, actinomycin D, doxorubicin, cyclophosphamide, ifosfamide and etoposide. Although vincristine, actinomycin D and cyclophosphamide is the gold standard combination chemotherapy regimen for rhabdomyosarcoma, other combinations are used. Patients with metastatic disease have a poor prognosis despite aggressive therapy.

Tumours of adipose tissue

Lipoma

Lipomas are benign tumours of adipose tissue and they are among the commonest of all benign neoplasms. They may occur in any organ as fat is present throughout the body but they are most often found in the subcutaneous tissue of the neck and shoulders, chest and thigh. The incidence of lipomas increases over the age of 40 and they are more commonly found in women than in men. Histologically, lipomas are usually simple but there are several other types that are defined by their histological appearance. Fibroblastic or spindle cell lipomas are made up of fat cells and fibroblasts in a myxomatous stroma. These are benign but may be difficult to distinguish from liposarcomas on histology. Angiolipomas are characterized by a dense vascular plexus.

Clinically, a lipoma presents as a slow-growing asymptomatic mass in the subcutaneous tissue. Occasionally, deeper lipomas may cause nerve compression and are then associated with paraesthesia, pain and other functional problems. On examination, they are typically circumscribed, non-tender and are characteristically mobile, as they are not attached to skin. Small lipomas may be difficult to palpate as they slip from under the examining fingers. Angiolipomas are characterized by pain, and *Dercum disease* (adiposa dolorosa) is a syndrome in which diffuse adipose tissue is deposited around the ankles, knees and elbows in menopausal women. This is often associated with pain.

Although most lipomas are subcutaneous, they may infiltrate fascia, muscle, periosteum or bone. They may recur after excision, but do not behave like malignant tumours. The treatment of a single lesion is excision in order to obtain histological diagnosis. Patients who have multiple lipomas should only have those lesions excised that are causing problems. If these are confirmed histologically as lipomas, it is safe to assume that the other lesions are similar.

Liposarcoma

These malignant tumours of adipose tissue are rare but must always be considered in the differential diagnosis when a lipoma is suspected. Unlike lipomas, liposarcomas usually arise from deeper tissues. They occur most frequently in older patients and are commoner in men than in women. The commonest sites are the limbs and the retroperitoneum. Histologically, liposarcomas are classified according to their malignant potential into five subtypes:

1. well differentiated
2. myxoid
3. fibroblastic
4. lipoblastic
5. pleomorphic.

The *well-differentiated* subtype is made up of mature adipocytes which show some pleomorphism and occasional mitotic figures. These tumours have a good prognosis. They do not metastasize but are invasive and tend to recur locally. The second subtype, *myxoid liposarcoma*, consists of pleomorphic adipocytes and preadipocytes and fibroblast-like cells or prolipoblasts. This subtype is commonly found in the groin, thigh, trunk, popliteal fossa and the retroperitoneum. It is most commonly seen in the fifth and sixth decades of life. These tumours may metastasize and have a 50% risk of local recurrence after excision. The *fibroblastic* subtype is made up largely of prolipoblasts with a few preadipocytes. It presents clinically very much as the myxoid liposarcoma but has a poor prognosis with a 62% recurrence rate after surgical excision. *Lipoblastic liposarcoma* is made up of round cells with preadipocytes and lipoblasts in a myxoid stroma. This tumour shows a similar malignant potential to the fibroblastic type. The *pleomorphic liposarcoma* is made up of undifferentiated cells with a high mitotic rate. The tissue of origin is demonstrated by the presence of occasional preadipocytes, lipoblasts or intracytoplasmic hyalin globules. This tumour has a strong tendency to metastasize and is associated with a very poor prognosis.

Treatment of liposarcoma is by wide excision, and for the less differentiated types radiotherapy should be considered either as adjuvant treatment or as an alternative when surgical excision is impossible or incomplete.

Tumours of the lymphatic system

Lymphangioma

Lymphangiomas are thought to arise from failure of the lymphatics to connect to the venous system, from abnormal budding of lymphatic tissue or from sequestered lymphatic rests during embryonic growth. There is some evidence that vascular endothelial growth factor C and its receptors may be involved in the development of lymphangiomas. Whatever the exact cause, the abnormal 'lymphatic rests' penetrate adjacent structures and fascial planes during embryonic development, and, as their lining retains secretory function, they form cystic lesions that do not communicate with the venous system. The nature of the tissues that the developing lymphangiomas permeate determines their eventual type: *capillary, cavernous or cystic*. Thus, cystic hygromas (see below) tend to occur in loose areolar tissues, whereas in muscle lymphangiomas tend to be of the capillary or cavernous types.

Giguere *et al*. have proposed categorization of lymphangiomas based on the size of the cystic component, as follows:

- macrocystic: cystic spaces of at least 2 cm
- microcystic: spaces <2 cm
- mixed lesions.

Lymphangioma simplex arises in the skin or mucosa and is common on the tongue. Several cystic lesions may occur together and this is called *lymphangioma circumscripta*. Treatment is usually by excision, although recurrence is relatively common. Electrocoagulation and argon laser therapy have also been used to good effect. The *cavernous lymphangiomas* are more extensive than lymphangioma simplex. They occur at any time between birth and about 20 months of age. They resemble cavernous haemangioma but lack the blue colour seen with this lesion. Initially, they are soft but after repeated infections they become fibrotic. Some may resolve spontaneously and if this occurs it happens usually before the age of 8 years. If they persist after this age then treatment should be by wide surgical excision if this is possible.

Lymphangiomas may also be acquired as a result of trauma, including surgery and inflammatory conditions leading to obstruction/destruction of the lymphatic channels. Acquired lymphangiomas give rise to swelling of the limb from lymphoedema.

Cystic hygroma

Cystic lymphangioma, more commonly referred to as cystic hygroma, is a cystic lymphatic lesion that may occur at any site, but 75% of lesions affect the head and neck with a left-sided predilection usually in the posterior triangle. The next most common site of cystic hygroma is the axilla (20%). Other less frequent sites include the mediastinum, groin and retroperitoneum. Morbidity depends on the anatomical location and is related to cosmetic disfigurement, bleeding into the cyst causing sudden enlargement, infection and compression of vital structures, e.g. airway, nerves and vessels.

Clinical features

The incidence of cystic hygroma is estimated to be one case per 6000–16 000 live births. Most cystic hygromas (50–65%) are present at birth, and the majority (80–90%) present within the first 2 years of life. Signs and symptoms of cystic hygroma vary, depending on location. Most commonly, cystic hygromas are located below the mylohyoid muscle either in the anterior or posterior triangles of the neck.

de Serres *et al*. have proposed the system shown in Table 16.7 for staging of cystic hygroma of the head and neck, which has practical importance in terms of associated morbidity.

Cystic hygromas are typically soft, painless, compressible masses which transilluminate vividly. At presentation, the cysts are typically large and thick walled with little involvement of surrounding tissues and a normal/slightly bluish tinged overlying skin. Although most cystic hygromas are multicystic, in 10% of cases they consist of a unilocular cyst. Cysts can range from 1 mm to several centimetres in size. They are filled with clear to straw-coloured protein fluid and are surrounded by a fibrofatty stroma that contains lymphoid aggregates or smooth muscle.

Cystic hygromas may enlarge and present after an upper respiratory or secondary infection and, more rarely, intracystic haemorrhage. Less commonly, children with cystic hygroma may display symptoms of obstructive sleep apnoea from lesions located in the supraglottic/paraglottic regions. These children may present acutely with stridor and cyanosis. Other symptoms include feeding difficulties and failure to thrive. Cystic hygroma is reported to be more common in patients with Turner, Down and Klinefelter syndromes, although these are not aetiologically related.

The imaging tests used in the diagnosis of cystic hygroma are ultrasonography and MRI. CT scanning should be avoided because of the radiation hazard in infants and children.

Treatment

Expectant treatment is indicated only in patients without symptoms and a static condition; otherwise, active treatment is necessary. The medical treatment of cystic hygroma consists of aspiration and instillation of sclerosing agents, e.g. OK-432 (Picibanil; an inactive strain of group A *Streptococcus pyogenes*), ethanol, sodium tetradecyl sulphate and doxycycline. OK-432 works best for large unilocular cysts. Fibrin sealant has also been reported recently to give good results

However, only surgical resection offers the potential for cure. In the absence of urgent symptoms, surgery can be delayed until the child is aged 2 years. Complete removal offers the best

Table 16.7 Staging of cystic hygroma of the head and neck (after de Serres *et al.*)

Stage	Complication rate
Stage I: unilateral infrahyoid	17%
Stage II: unilateral suprahyoid	41%
Stage III: unilateral and both infrahyoid and suprahyoid	67%
Stage IV: bilateral suprahyoid	80%
Stage V: bilateral infrahyoid and suprahyoid	100%

chance of cure. If this is not possible because of risk of damage to vital structures, as much as possible of the lesion is removed, accepting an increased risk of recurrence. Complications of surgical excision of cystic hygroma include damage to cranial nerves, chylous fistula, chylothorax, haemorrhage and recurrence. Most recurrences occur within the first year.

Lymphangiosarcoma

Lymphangiosarcoma is a rare (fewer than 250 cases reported in the literature) malignant neoplasm arising from the endothelium of blood vessels or lymphatics in patients with longstanding congenital or acquired lymphoedema. The vast majority of cases have been reported in association with postmastectomy lymphoedema, when is usually referred to as Stewart–Treves syndrome. Rarely, the disease may develop in patients with congenital hereditary and non-hereditary lymphoedema. Hereditary lymphoedema is of two distinct types: the lymphoedema is present at birth or develops in early childhood (Milroy disease), and late onset disease, which may develop between the first decade of life and late puberty (Meige syndrome). Very rarely, lymphangiosarcoma may develop in patients with non-hereditary types of lymphoedema including filariasis.

In the Stewart–Treves syndrome, lymphangiosarcoma develops in the grossly oedematous upper limb some 10–20 years after radical mastectomy. The tumour forms purplish papules or nodules which may coalesce or undergo necrosis and ulceration. The prognosis is poor because of the tendency to both local recurrence and distant metastases, especially to the lungs. The treatment consists of neoadjuvant radiotherapy followed by ablative surgery. Lymphangiosarcoma is not responsive to chemotherapy. The overall reported 5 year survival is 20%.

Vascular tumours

The majority of vascular tumours are benign. Previous classifications were based on their clinical appearance, the architecture of the vascular channels and the embryology. However, Mullikan has introduced a classification based on cellular activity, which has been generally accepted because of its important clinical implications. Vascular abnormalities can be divided into true haemangiomas and vascular malformations. The term *haemangioma* should be restricted to those lesions which grow by proliferation of endothelial cells. Clinically, they are absent at birth and tend to grow rapidly in a proliferative phase for the first year of life. During this time their growth is disproportionately high with respect to the child's growth. The lesions then enter an involutional phase during which they regress spontaneously. *Pyogenic granulomas* are included in this category. *Vascular malformations* on the other hand are structural abnormalities that result from embryonic defects in morphogenesis. These grow in parallel with the patient's growth and do not regress. Many lesions which were formally termed cavernous haemangiomas fall into this group.

Haemangiomas

Haemangiomas constitute the commonest tumour of childhood and affect 10–12% of white children at 1 year old. They occur predominantly in the head and neck but are also found on the trunk and limbs (Figure 16.15). The diagnosis is based on clinical grounds and relies on a history of a lesion arising soon after birth and growing initially at a rate higher than the child's growth.

A superficial haemangioma or *strawberry naevus* is bright red and raised with a characteristic appearance. As it begins to involute, generally in the second year of life, it develops a pitted appearance. Deeper lesions may appear blue through the overlying skin. Because of rapid enlargement, haemangiomas may give rise to a number of complications. About 5% ulcerate and a few develop infection. The resulting tissue destruction can be severe and these are referred to as wildfire haemangiomas.

Occasionally the anatomical position of a haemangioma may cause problems, e.g. haemangiomas of the respiratory or alimentary tracts may give rise to obstruction and haemangiomas on the eyelid may result in amblyopia and failure to develop binocular vision. Bleeding may occur locally or, rarely, the patient may develop platelet thrombi within the haemangioma, leading to a consumptive thrombocytopenia (*Kasabach–Meritt syndrome*). These patients require steroid therapy. In general, the treatment of haemangioma should be expectant, although occasionally, when there is severe bleeding or infection, excision may be required.

Pyogenic granuloma (lobular capillary haemangioma)

Inaptly named 'pyogenic granuloma', nowadays this condition is referred to as lobular capillary haemangioma as it is neither infectious nor granulomatous in origin, although its aetiology remains unknown. Instead it consists of a localized proliferative vascular lesion common in children and young adults, typically

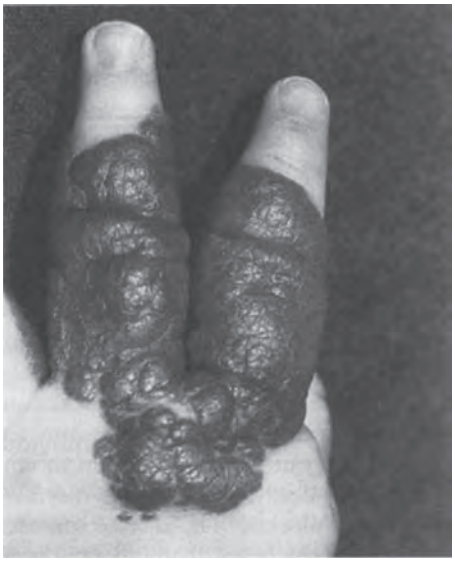


Figure 16.15 Haemangioma.

found on the face or limbs. There may be a history of trauma but usually it arises spontaneously and forms a cherry-red pedunculated lesion that often ulcerates and may become infected or bleed. The condition may also arise in pregnancy (pregnancy tumour) and, less commonly, with oral contraception either in the gingiva or elsewhere in the oral mucosa. As with other types of haemangioma it may involute if treated expectantly, but, if bleeding or infection occurs, the lesion may be treated by silver nitrate cautery, electrocoagulation or surgical excision.

Vascular malformations

As indicated above, vascular malformations (hamartomas) are structural abnormalities which are determined during embryological development. This may result in a capillary, venous, lymphatic or combined abnormality. Any of these can produce a low-flow malformation. In some cases, however, an arteriovenous fistula may develop, leading to a high-flow malformation.

Port-wine stain

The port-wine stain is a low-flow capillary malformation, sometimes called a capillary haemangioma or *naevus flammeus*. It presents at or soon after birth as a flat area of discoloration. Initially it is pink or scarlet but with age it deepens to purple. It grows in parallel with the patient's growth, but with increasing age it may become macular, giving a cobblestone type of appearance. On histological examination, it is made up of mature vascular endothelium. Port-wine stains on the trunk and limbs may be associated with lymphatic or venous abnormalities (*Klippel Trénaunay syndrome*). Port-wine stains of the face may be associated with underlying abnormalities of the meningeal and choroid plexus vasculature, resulting in focal seizures (*Sturge–Weber syndrome*). Port-wine stains are best treated using photocoagulation with a tuneable dye laser. This lightens the colour of the lesion and reduces the cobblestone effect in older lesions. Very rarely, excision and skin grafting may be necessary.

Capillary lymphatic malformations

These lesions, also known as *lymphangioma circumscriptum* or *verrucous haemangiomas*, are commonly found on the limbs and trunk. They present at birth as clearly demarcated raised red areas. With age they become hyperkeratotic, giving them a warty appearance. On histology they are made up of dermal and subcutaneous vessels and lymphatics. Treatment, when necessary, is by surgical excision and it is important to remove any deep extensions of the lesion to prevent recurrence.

Angiokeratoma

Angiokeratomata are abnormalities of the microvasculature presenting as dark red/purple spots up to 1.0 cm in diameter and sometimes with a rough scaly surface and composed of dilated capillaries. There are four types: (1) sporadic, (2) angiokeratoma of Fordyce, (3) angiokeratoma circumscriptum and (4) Fabry syndrome (*angiokeratoma corpus diffusum*). The sporadic type (angiokeratomata of Mibelli) forms solitary lesions on the hands and feet in patients over 40 years old. The angiokeratomata of Fordyce are found in linear groups most commonly on the scrotum, but also on the shaft of the penis, labia major, vulva, inner thigh and lower abdomen. Angiokeratoma circumscriptum often presents as a birth mark but may develop later in life. It is commoner in females and usually consists of a localized cluster of lesions in the leg or trunk. Fabry disease is a rare, serious, inherited sex-linked recessive disorder of glycolipid metabolism (deficiency of -galactosidase A), which results in abnormal deposits of glycophospholipids in the blood vessels and internal organs. Aside from multiple angiokeratoma, these patients often develop fever and painful hands and feet (acroparaesthesias) and corneal and lenticular changes. They are prone to develop renal failure, strokes and hypertensive cardiovascular disease.

Telangiectasia

The telangiectasia lesion is known as a spider naevus and consists of a dilated vessel with radiating branches. It is commonly found on the face, arms and chest and, because of its vascular nature, it blanches on pressure over the central feeding vessel. Spider naevi are found in pregnancy and in liver cirrhosis, in which they are thought to be related to high oestrogen levels. Hereditary haemorrhagic telangiectasia (Osler–Weber–Rendu disease) is characterized by multiple telangiectasia and is complicated by widespread bleeding and the late development of arteriovenous pulmonary malformations. Individual spider naevi rarely require treatment but, if removal is required for cosmetic reason, thermocoagulation may be used.

Venous malformations

Venous malformations are also known as cavernous haemangiomas. Clinically they are soft, compressible and non-pulsatile and may present in any site. They are often dark blue, but in deep tissue when they are covered by skin they may have a normal colour. If combined with a capillary malformation they may be dark red. The adjacent bone may become distorted or hypertrophied in response to the blood flow through the lesion. Phlebothrombosis may develop within the abnormality, giving rise to pain and induration of the lesion. Minor injury may lead to rapid enlargement of the vessels or the formation of the arteriovenous fistula. Treatment of venous malformations must be planned by delineating the extent of the abnormal vessels. Angiography can confirm the venous nature of the lesion and define feeding vessels, and MRI may be useful to define the extent of the lesion, which is not seen on angiography. Treatment depends on the size of the lesion and its anatomical site, but may consist of intravascular coagulation, selective arterial embolization of feeding vessels or surgical resection.

Arteriovascular malformations

Compared with low-flow abnormalities, high-flow vascular malformations are rare. They tend to occur on the head and neck and limbs and may arise from an innocent low-flow capillary vascular malformation. Trauma or the hormonal changes of puberty and pregnancy may precipitate the formation of arteriovenous communications in such a lesion. Clinically they present as a raised, warm lesion with a palpable thrill and a bruit on auscultation. The overlying skin may become ischaemic and the lesion may haemorrhage spontaneously or as a result of relatively minor trauma. The high flow may cause gigantism of the limb or the digit and the pulsatile flow may erode bony structures. There may also be a systemic effect as the shunt effect of the fistula may give rise to cardiac failure. Treatment of an arteriovascular malformation generally involves surgical excision, although preoperative embolization may help to reduce intraoperative blood loss.

Malignant vascular tumours

Haemangiopericytoma

Haemangiopericytoma is a rare tumour that usually occurs in adult life (fifth decade) and less commonly in children (10% of cases). Haemangiopericytoma arises from pluripotential cells (pericytes) in the walls of capillaries. Multiple molecular genetic changes are involved in the oncogenesis of haemangiopericytoma, including activation of certain oncogenes of the insulin-like growth factor family.

Clinically, haemangiopericytoma forms a deep soft-tissue tumour (usually muscle, rarely subcutaneous or dermal), which may occur anywhere including the brain, but the most common sites are the lower limbs, pelvis and head and neck regions. Clinical presentation is with a slowly enlarging painless mass, with pain becoming a feature only in advanced disease. Rarely haemangiopericytoma may present with hypoglycaemia owing to the inappropriate secretion by the tumour of an insulin-like growth factor. Prognosis is worse in adults with haemangiopericytoma, with the development of systemic metastases (commonly lungs and bones) and a 5 year survival of 50%.

Treatment

Haemangiopericytoma is treated by complete surgical excision, which remains the mainstay of treatment. Preoperative ligation or vascular embolization of the afferent vessels is indicated for large tumours, largely to reduce blood loss during removal. Chemotherapy and radiotherapy are indicated in patients in whom complete resection is not possible and in large locally invasive tumours. Adjuvant postoperative radiotherapy (50 Gy) is recommended even in patients with complete excision in view of the reported improved local control.

Kaposi's sarcoma

Kaposi's sarcoma (KS), first described in 1872 by the Hungarian dermatologist Moritz Kaposi, used to be a rare vascular tumour affecting elderly men of Italian or Eastern European Jewish ancestry. However, nowadays, the neoplasm is encountered in several other populations – young and adult black African males, patients on long-term immune suppression after allograft transplantation and homosexual HIV-negative males – in addition to the fulminant form of KS, associated with HIV disease and referred to as epidemic KS. Although the histopathology of the different clinical types in these various patient groups is essentially the same, the clinical manifestations, course of the disease and prognosis differ considerably from one group to the other. The human herpes gamma virus type 8 (HHV-8), also known as KS herpes virus (KSHV), is present in all biopsies of KS irrespective of type and is absent in non-involved tissue, indicating an aetiological role for HHV-8.

Kaposi's sarcoma-associated herpesvirus

In view of its aetiological importance to the development of KS and, indeed, other lymphoproliferative disorders, it is necessary to outline the nature, epidemiology, transmission and pathological consequences of human infection with this virus. It was first isolated by Chang *et al.* in 1994 from a KS skin lesion from a patient with AIDS. The virus is now referred to as KSHV or HHV-8. Since then, it has been documented that virtually all KS lesions, irrespective of clinical subtype, are infected with KSHV. Hence, there is little doubt that this virus is the primary factor in the aetiology of KS. However, an important cofactor in the development of clinical KS is immunosuppression of the host. In this sense, therefore, KS differs from other neoplasms in that reactivation of KSHV in the immunocompromised host is necessary for its development. It remains uncertain whether the KS lesions are composed of neoplastic cells harbouring specific cytogenetic alterations (mutations, rearrangements and amplifications) or are, in essence, the result of HHV-8-induced abnormal intracellular signalling pathways that modulate the expression of cellular genes associated with cell cycle regulation, apoptosis, inflammatory response and angiogenesis. According to this viewpoint, these are not true tumours but represent a reactive angioproliferative disorder. KSHV is also associated with the development of other lymphoproliferative disorders, e.g. body cavity-related -cell lymphoma and some plasma cell forms of multicentric Castleman disease.

KSHV is transmitted by sexual and other means. Sexual transmission is more common with homosexual (anal recipients) than heterosexual contact, and the prevalence of KSHV infection increases with the number of male sexual partners. Other modes of transmission predominate in African countries, where infection can occur during childhood through maternal–infant transmission (during labour or delivery or transplacental). Other non-sexual routes of transmission operate in children; the exact modes are not known, although KSHV has been detected in the saliva of infected persons. Clinical disease in transplant patients results predominantly from reactivation of the virus. One study on 220 transplant recipients reported antibodies against KSHV (seroconversion) in 25 patients within the first year after transplantation. The risk of transmission of KSHV through blood products is not known.

The KSHV infection rates parallel the incidence of KS, being low in the USA, many parts of Europe and Asia, intermediate in Mediterranean countries and highest in Central Africa. The seroprevalence of KSHV among blood donors ranges from 10% in the USA to >50% in many African countries, with intermediate rates in Italy and other Mediterranean countries.

Most primary KSHV infections are asymptomatic since healthy immunocompetent adults are able to suppress the KSHV infection, whereas the infection becomes clinically symptomatic in the presence of immunodeficiency. In HIV seropositive patients, the incubation period for disorders caused by KSHV infection depends on the immune status of the patient. In patients with AIDS infected with KSHV, the ability of human leucocyte antigen class I-restricted cytotoxic T-lymphocytes to respond to KSHV proteins becomes lost as the immunodeficiency worsens, at which point KS develops. Both KS and -cell body cavity lymphoma respond dramatically to active antiretroviral therapy in patients with AIDS. Additionally, KS in transplanted patients often regresses when immunosuppressive regimens are discontinued or reduced.

Clinical types of Kaposi's sarcoma

The types or variants of KS that are recognized are shown in Table 16.8.

Classic Kaposi's sarcoma

This is a rare disease with a predilection for males (reported male to female ratio of 15:1). The usual age at onset is between 50 and 70 years, the disease presenting with one or more firm asymptomatic red, purple or brown patches, plaques or nodular skin lesions. Classic KS is often localized to one or both lower limbs, initially around the ankles and soles of the feet but gradually progresses up the arms and legs. On histology, the tumour is composed of spindle-shaped cells surrounding hyperaemic vascular clefts, often in association with extravasated erythrocytes, haemosiderin and fibrosis. Classic KS runs a relatively benign, indolent course for several years with slow enlargement and gradual emergence of new lesions. Venous stasis and lymphoedema may precede or follow the appearance of the skin lesions. Systemic lesions (usually in the gastrointestinal tract) may develop in longstanding cases. These visceral lesions are usually asymptomatic although occasionally they may cause gastrointestinal bleeding. Some published series have reported the development of a second primary malignancy, usually a body cavity lymphoma.

Treatment

In patients who are immunocompetent and have single or few localized lesions, treatment is by surgical excision as this provides an adequate outcome but requires prolonged follow-up because of the risk of recurrence (40%). In contrast multifocal disease is unsuitable for surgical treatment in view of high recurrence rates. Patients with several lesions limited to a localized area are best treated with single doses of radiotherapy (8–12 Gy) as this treatment modality gives good control with reported symptomatic relief in the majority, partial regression in 85% and a complete response in 50%.

Patients with extensive or recurrent KS are treated either with a combination of surgery and chemoradiotherapy or with chemotherapy alone as this can result in complete remission lasting several years. The active cytotoxic agents used are vinblastine, bleomycin, doxorubicin, and dacarbazine alone or in combination. Intralesional injection of interferon 2b in a dose of 1×10^6 to 3×10^6 units is an alternative in view of reported good response rates. Its advantage over chemotherapy and radiotherapy is the avoidance of systemic toxicity.

Table 16.8 Types or variants of Kaposi's sarcoma

Clinical type	Population at risk	Prognosis
Classic	Elderly males of Eastern European/Mediterranean origin	Long survival
Endemic	African children and adults	Survival months to years
Immunosuppression	Organ transplant recipients	Survival months to years
Epidemic/AIDS associated	HIV-infected patients – homosexual and bisexual males	Survival weeks/months

Endemic (African) Kaposi's sarcoma

First reported in 1950, KS is a common neoplasm endemic in native populations in equatorial Africa. Nowadays, two types are recognized: endemic KS in HIV-negative patients and KS in HIV-positive patients. Together, these two types account for most of the soft-tissue tumours in central Africa. Adult endemic KS may occur either as an indolent neoplasm identical to the classic disease with nodules or plaques in oedematous limbs or as an aggressive fungating and exophytic lesion that invades the subcutaneous tissue and bone. Both types are again much more common in males but present at significantly lower age group (median age 45) than the classic disease, and a small percentage (8%) have concurrent lymphoma. A third type, the lymphadenopathic form of KS seen in Africa, occurs usually in prepubescent children. It is characterized by generalized lymphadenopathy, which is often associated with visceral organ involvement and a very poor prognosis with death in most patients within 3 years of diagnosis.

Treatment

Endemic KS is treated by chemotherapy using the same cytotoxic agents used in classic disease with reported response rates of up to 80%.

Immunosuppressive treatment-related Kaposi's sarcoma

This is encountered in organ transplant patients on immunosuppressive therapy in whom the incidence is estimated at 150–200 times that of the general population. The French Collaborative Transplantation Research Group reported a 0.5% overall risk of KS in 7923 organ transplant recipients. The male predominance persists but is much less marked than in classic KS, with reported male to female ratios ranging from 2:1 to 4:1. The average time to develop KS after transplantation is 29–31 months (range 3–124 months). In general, this type of KS tends to be aggressive, involving lymph nodes, mucosa and visceral organs in about 50% of patients. Some patients have concurrent lymphoma, tuberculosis or transfusion-related HIV infection. These KS tumours may regress following reduction or changes in immunosuppressive therapy. Clinical management of transplant patients who develop KS is difficult and requires a balance between the risk of death from generalized KS and the risk of graft rejection if the immunosuppressive therapy is discontinued.

Treatment

The options in transplant patients who develop KS include (1) withdrawal or reduction of immunosuppressive therapy, (2) radiotherapy and (3) chemotherapy. In most studies the withdrawal/reduction of immune suppression results in graft rejection in approximately 50%. Total withdrawal, which often results in complete response, is an acceptable risk only in renal transplant patients in view of the availability of dialysis (in the event of graft rejection), but obviously not in patients with cardiac or liver transplants in whom the immunosuppressive drugs can only be reduced or modified. In these patients, chemotherapy, single or combination with the cytotoxic agents used for classic KS (doxorubicin, bleomycin and vincristine) with or without radiotherapy, is used. There is a good reason to institute chemoprevention in transplant patients who are seropositive for KSHV. The problem is that the current antiviral drugs that are active against KSHV (ganciclovir, foscarnet and cidofovir) are too toxic for long-term chemoprevention.

Epidemic Kaposi's sarcoma

The proportion of HIV disease patients with KS has steadily decreased since the AIDS epidemic was first identified in 1981. About 48% of AIDS patients in 1981 had KS as their presenting AIDS diagnosis. By 1987, this had decreased to less than 20%. This is attributed to the introduction of highly active antiretroviral therapy (HAART), which delayed or prevented the emergence of drug-resistant HIV strains, thereby substantially reducing the viral load – the risk of opportunistic infections accounting for the decline of KS incidence in AIDS patients. In epidemic KS, the lesions may involve the skin, oral mucosa, lymph nodes and visceral organs (gastrointestinal tract, lung, liver and spleen). Eventually, most patients with epidemic KS develop disseminated disease. The disease often progresses rapidly from a few localized or mucocutaneous lesions to more numerous lesions and generalized skin disease with lymph node, gastrointestinal tract disease and other organ involvement. Pleuropulmonary KS is an ominous sign and usually cause death directly attributed to KS. Also, many patients with epidemic KS die of one or more complicating opportunistic infections. Several reports have documented KS in homosexual HIV-negative men. These patients have an indolent and cutaneous form of the disease, which causes new lesions to appear every few years. The lesions occur most commonly on the extremities and genitalia but can occur elsewhere in the skin.

Treatment

In AIDS patients who present with KS, the initial treatment is with HAART. If the patient responds to this with restoration of immune competence, this will have a beneficial effect on the KS lesions and certainly halt the progression of the disease. A good response to HAART may induce shrinkage of the KS lesions and reduce the KSHV viral load. However, not all patients respond to HAART; this varies considerably from patient to patient. For this reason, specific local or systemic therapy for KS is usually instituted as well.

Radiotherapy is used for lesions limited to areas of the skin or oral mucosa. Systemic chemotherapy is indicated for patients with widespread mucocutaneous disease, lymphoedema or visceral disease. The most active drugs include liposomal anthracyclines, paclitaxel, vinca alkaloids and bleomycin. Interferon- (subcutaneous, intralesional or intravenous) is also effective for epidemic cutaneous KS. The best results with this biological agent (when used singly) are reported with daily intramuscular or intravenous injections of interferon- in a dose of 30×10^6 units. The common side effects include fever, chills, fatigue and muscle pain. One RCT has shown that the combination of zidovudine and interferon- (1×10^6 units subcutaneously per day) produces a higher response rate (31% versus 8%) than interferon- alone.

A more recent therapeutic approach for epidemic KS is with use of inhibitors of angiogenesis (vascular endothelial growth factor and thalidomide) and retinoids (all-*trans* retinoic acid). Early results are promising but this therapy remains experimental.

GUIDE TO FURTHER READING

Antman K, Chang Y. Kaposi's sarcoma. *N Engl J Med* 2000;**342**:1027–1038.

Boshoff C, Schulz TF, Kennedy MM, *et al.* Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. *Nat Med* 1995;**1**:1274–1278.

Crist WM, Anderson JR, Meza JL, et al. Intergroup Rhabdomyosarcoma Study-IV: results for patients with nonmetastatic disease. *J Clin Oncol* 2001;**19**:3091–3102.

Gibbs S, Harvey I, Sterling J, Stark R. Local treatments for cutaneous warts: systematic review. *BMJ* 2002;**325**:461.

Hallberg Ö, Johansson O. Malignant melanoma of the skin – not a sunshine story! *Med Sci Monit* 2004;**10**: CR336–340.

Koops HS, Vaglini M, Suciu S, *et al.* Prophylactic isolated limb perfusion for localized, high-risk limb melanoma: results of a multicenter randomized phase III trial. European Organization for Research and Treatment of Cancer Malignant Melanoma Cooperative Group Protocol 18832, the World Health Organization Melanoma Program Trial 15, and the North American Perfusion Group Southwest Oncology Group-8593. *J Clin Oncol* 1998;**16**:2906–2912.

Lee JY-Y, Yang C-C, Chao S-C. Histopathological differential diagnosis of keloid and hypertrophic scar. *Am J Dermatopathol* 2004;**26**:379–384.

Meirion J, Thomas JM, Newton-Bishop J, *et al.* Excision margins in high-risk malignant melanoma. *N Engl J Med* 2004;**350**:757–766.

Schrama D, Keller G, Houben R, *et al.* $\text{BRAF}^{\text{V600E}}$ mutations in malignant melanoma are associated with increased expressions of BAALC. *J Carcinog* 2008;**7**:1.

Stewart FW, Treves N. Lymphangiosarcoma in postmastectomy lymphedema: a report of six cases in elephantiasis chirurgica. *Cancer* 1948;**1**:64–81.

Wong CSM, Strange RC, Lear JT. Basal cell carcinoma. *BMJ* 2003;**327**:794–798.

Wood NH, Feller L. The malignant potential of HIV-associated Kaposi sarcoma. *Cancer Cell Int* 2008;**8**:14.

Yang J, Chang AE, Baker AR, *et al*. Randomized prospective study of the benefit of adjuvant radiation therapy in the treatment of soft tissue sarcomas of the extremity. *J Clin Oncol* 1998;**16**:197–203.

CHAPTER 17

Disorders of the breast

ALASTAIR M. THOMPSON, STACY L. MOULDER-THOMPSON, ANDREW EVANS AND JOHN A. DEWAR

Introduction

Many women develop breast symptoms during their lifetime, most of which are benign, self-limiting and resolve in time. Some symptoms should be referred to a breast specialist (Box 17.1) whereas others can be managed in the community (Box 17.2).

Breast conditions can be subdivided into benign and malignant diseases. Whatever the underlying pathological process, a concise history and examination followed by appropriate imaging and pathological investigation is required.

Organization of services

Ideally, breast clinics should be established and served by staff trained in breast disease: surgeons, breast physicians, radiologists, diagnostic radiographers and breast care nurses, with specialist pathology, oncology and reconstuctive surgery back-up.

Patients attending for diagnostic purposes should be seen by a clinician trained in the management of breast conditions. Based on evidence that patients with breast cancer managed in a breast unit/centre have a better outcome, patients with breast cancer should be managed by a multidisciplinary team within a designated breast unit.

While delay in being seen at a breast clinic is associated with marked anxiety, delays in onset of treatment of <3 months are unlikely to be associated with a measurable difference in survival.

BOX 17.1 Patient referral: conditions that require referral to a breast specialist

- Lump
  - Any new discrete lump
  - New lump in pre-existing nodularity
  - Asymmetrical nodularity that persists at review after menstruation
  - Abscess or breast inflammation which does not settle after one course of antibiotics
  - Cyst persistently refilling or recurrent cyst (if the patient has recurrent multiple cysts and General Medical Practitioner has the necessary skills, then aspiration is acceptable)
- Pain
  - If associated with a lump
  - Unilateral persistent pain in postmenopausal women
- Nipple discharge
- All women aged 50 and over
- Women under 50 with:
  - Blood-stained discharge; or
  - Bilateral discharge sufficient to stain clothes; or
  - Persistent single duct discharge
- Nipple retraction or distortion, nipple eczema
- Change in skin contour

BOX 17.2 Patient referral: conditions that can be initially managed in general practice

- Young women (<35 years) with tender, lumpy breasts and older women with symmetrical nodularity, provided that they have no localized abnormality
- Women with minor and moderate degrees of breast pain who do not have a discrete palpable lesion
- Women aged under 50 years who have nipple discharge that is from more than one duct or is intermittent and is neither bloodstained nor troublesome

Hence, in the UK, it has been suggested that >80% of urgent referrals are seen within 5 working days and the remainder within 10 working days (after receipt of referral).

History

The history for a patient with breast symptoms can be subdivided into:

- history of the presenting problem
- hormonal history
- gynaecological history
- family history
- other medical/surgical history
- breast imaging history.

History of the presenting problem

Having established the age and the sex of the patient, the answers to the questions shown in Box 17.3 should be recorded.

The categories shown in Box 17.4 identify women who have three times (or more) the population risk of developing breast cancer.

Hormonal and gynaecological history

This should include age at menarche and (when appropriate) menopause, number and timing of pregnancies, use of oral contraception and hormone replacement therapy (and whether oestrogen and/or progesterone containing, duration and timing of usage).

BOX 17.3 History for breast conditions

- Is there a lump?
- Where is the lump?
- How big is it?
- When was it first noted?
- Is it a single/are they multiple lump(s)?
- Have there been any previous lumps?
- Does it change with menses?
- Is it tender/mobile/hard/soft?
- Are there any lumps elsewhere?
- Are there any associated features?
- Skin changes
- Nipple indrawing
- Nipple discharge – single or multiple ducts; blood stained or not
- Are there any problems in the other breast, axilla, supraclavicular fossae or neck?

BOX 17.4 Familial breast cancer: criteria for identifying women who have three times (or more) the population risk of developing breast cancer

A woman who has

- one first-degree relative with bilateral breast cancer or breast and ovarian cancer, or
- one first-degree relative with breast cancer diagnosed under the age of 40 or one first-degree male relative with breast cancer diagnosed at any age, or
- two first- or second-degree relatives with breast cancer diagnosed under the age of 60 or ovarian cancer at any age on the same side of the family, or
- three first- or second-degree relatives with breast or ovarian cancer on the same side of the family

In this context, a first-degree female relative is mother, sister or daughter. A second-degree female relative is grandmother, granddaughter, aunt or niece.

Family history

A history of cancers diagnosed in the wider family and the age at diagnoses and type of cancer in each family member should be ascertained. Breast, ovarian, prostate and colon cancers are of particular relevance in familial predisposition to breast cancer.

Other medical/surgical history

Here the aim is to assess the patient's general medical, mental and social health, especially as regards fitness for surgery, possible radiotherapy or chemotherapy and breast reconstruction. In particular, it is essential to know about medications, whether the patient smokes and any history of venous thrombosis. A history of previous radiation therapy to the chest (e.g. for lymphoma) should be recorded.

The patient's age (Figure 17.1) and the history will indicate the likely diagnosis but a systematic breast examination is necessary to pick up additional features.

Breast imaging history

Knowing when patients last had breast imaging will influence the choice of imaging required. Unless there is serious clinical suspicion of malignancy, repeating a mammogram within 12 months of a previous mammography is often not necessary.

Examination

Examination includes the key steps outlined below.

The patient (naked to the waist) is examined in the presence of a chaperone.

With the patient sitting, the breast should be inspected for changes in contour due to skin or nipple tethering, which may be emphasized by asking the patient to raise her arms above her head and lean slightly forward. Tethering of a lesion to the chest wall can be assessed by asking the patient to put her hands on

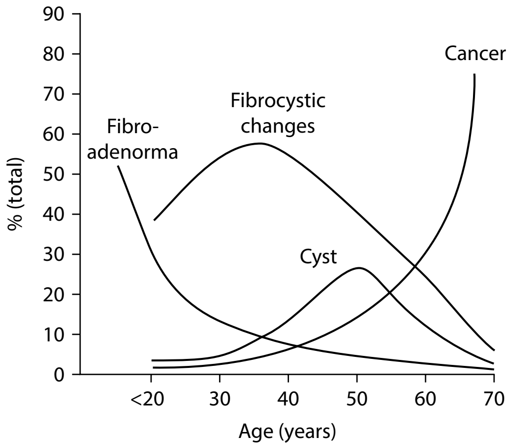


Figure 17.1 Incidence of breast lesions in women presenting to a breast clinic: fibroadenoma, fibrocystic changes, cyst and cancer as a percentage of breast lumps in patients presenting to a breast clinic compared with patient age.

her hips and press onto both hips, thus contracting the pectoral muscles. Next, with the patient lying comfortably with her hands behind her head, the normal breast should be examined first. Each breast should be examined, including the axillary tail, with the flattened fingers of the hand.

Any lump should be assessed for size (measured with calipers), shape, consistency, tenderness, mobility and fixity to the overlying skin/underlying chest wall.

Both axillae should be examined thoroughly for palpable nodes ensuring that the anterior, medial, posterior and apical node groups are palpated. The size, number and fixity of nodes should be recorded. Although clinical assessment of axillary nodes may not reflect their pathological status, for patients with breast cancer the clinical tumour–node–metastasis (TNM) system requires an assessment of node involvement.

The supraclavicular and infraclavicular fossae and cervical region should also be examined for palpable lymph nodes.

Any other symptomatic regions should be examined as required for signs (in a patient with breast cancer) of metastatic disease.

The examination findings must be recorded in sketch or written form, as part of good clinical practice and for medicolegal reasons.

Imaging

Most symptomatic problems in women under 40 who require imaging are best imaged using ultrasound. Ultrasound is excellent at discriminating cysts from solid lesions. Ultrasound is also highly accurate at benign/malignant differentiation of solid lesions. However, virtually all symptomatic solid lesions in women over the age of 25 years require a histological diagnosis. Most women attending a clinic over the age of 40 should have mammography unless they have had a mammogram within the last 12 months. All women with clinical findings suspicious of malignancy require mammography. Women with breast lesions thought to be malignant require axillary ultrasound with sampling of nodes abnormal on ultrasound. Most palpable lesions are best sampled under ultrasound guidance, which is more accurate than clinically guided biopsies.

Initial assessment, investigation and staging of breast symptoms

Methods of assessment of a breast abnormality include clinical history and examination, imaging and core biopsy sampling of the lesion for histological assessment. These three investigations collectively constitute *triple assessment*. There is strong evidence that triple assessment provides more accurate diagnosis than fewer tests. Clinical examination, mammography, ultrasound and core biopsy range in sensitivity to detect malignancy from 85% to 95%.

All patients should have a full clinical examination. When a localized abnormality is present, patients should have imaging usually followed by core biopsy. A lesion considered malignant on either clinical examination or imaging alone should have histopathological confirmation of malignancy before any definitive surgical procedure, e.g. mastectomy or axillary clearance. Clearly there are medicolegal implications if appropriate assessment is not performed prior to mastectomy if the subsequent definitive pathology turns out to be non-malignant.

Benign breast conditions

Aberrations of normal development and involution

The cyclical variations in oestrogen and progesterone result in increased mitosis of breast epithelial cells around days 22–24 of the menstrual cycle, but apoptosis restores the tissue balance across the cycle.

Benign proliferations in the breast are often considered as aberrations of normal development and involution.

Localized benign nodularity

Localized benign nodularity, previously termed fibroadenosis or fibrocystic disease, occurs commonly in young women, often with bilateral nodularity that is usually accentuated premenstrually. Any focal nodularity should disperse after two menses (approximately 6 weeks). Lumpy, normal breast tissue is usually managed in general practice following repeat examination, with reassurance, but any persistent localized abnormality requires referral for triple assessment. Many benign masses disappear completely spontaneously over time.

Fibroadenoma (Figure 17.2)

There is a spectrum of aberrations of normal development and involution that ranges from a single fibroadenoma (termed a juvenile fibroadenoma in teenage girls) through giant fibroadenoma to phylloides tumour. A fibroadenoma forms from a breast lobule, a smooth, non-tender, mobile lump, usually in a young woman. It may be single, lobulated or multiple and does not show any changes in size with menstruation.

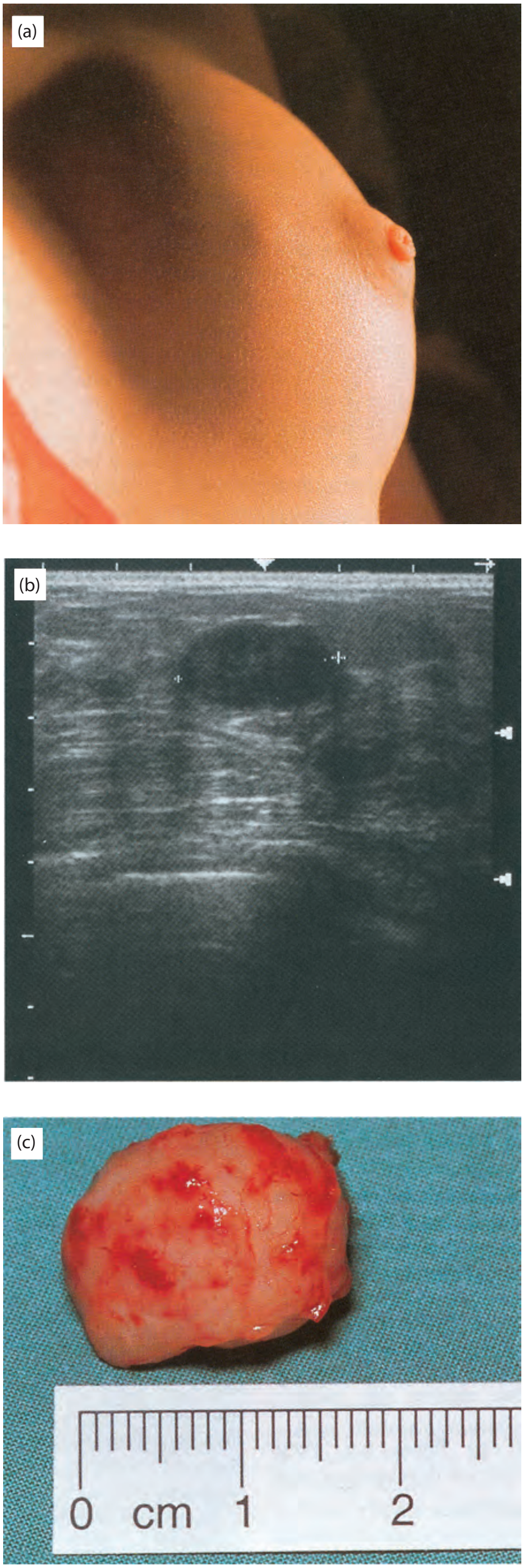


Figure 17.2 Fibroadenoma: (a) clinical appearance; (b) ultrasound appearance as a homogeneous, well-defined, oval mass; (c) surgical specimen.

Following diagnostic triple assessment the fibroadenoma can be excised if >4 cm or at the patient's request. Percutaneous vacuum biopsy represents an alternative to surgery for the removal of fibroadenomas <3 cm in size. Fibroadenomas are best sampled using core biopsy rather than cytology as, occasionally, papillomas and phylloides tumours may mimic fibroadenomas and these lesions are best diagnosed on core rather than cytology. A giant fibroadenoma should be excised as its increasing size may distort the shape of the breast. Phylloides tumours progressively increase in size, and range between a benign appearance and more sarcoma-like appearance with an increasing number of mitotic figures on histology. Local recurrence is common, if excision is incomplete, and therefore a phylloides tumour should be excised with a >1 mm margin of normal tissue to prevent local recurrence; even then further recurrence may require mastectomy; metastasis is very rare.

Cyst (Figure 17.3)

Rare before the age of 30 years, the incidence of cysts declines after the menopause. However, 7% of women develop a symptomatic cyst at some point during their life. Cysts may be single, multiple, unilateral or bilateral and range in size from 2 mm to several centimetres. Whereas ultrasound may be particularly useful in demonstrating the site and number of cysts, aspiration yields clear or turbid fluid ranging in colour from white to dark green; a cyst may disperse spontaneously; approximately 10% of symptomatic cysts recur and require reaspiration. Worrisome cysts are those that contain fresh blood, those in which there is a residual mass after aspiration and those that have a solid filling defect on ultrasound. Residual masses or intracystic filling defects should undergo core biopsy. Cysts which recur repeatedly also require histological assessment as a carcinoma in the wall of the cyst may be present. Endocrine agents such as danazol have been used to reduce the formation of further cysts in those prone to multiple cyst formation.

Duct papilloma

A duct papilloma classically presents as a blood-stained nipple discharge from a single duct; pressure on the areolar margin overlying the duct papilloma results in expression of a blood-stained discharge. The discharge can be tested for blood using a urinary testing stick. Papillomas can sometimes be seen on imaging as intraductal masses and then biopsied or even removed using a vacuum biopsy device under imaging control. Under general anaesthetic a lachrymal probe should be inserted into the duct and the single duct excised through a circumareolar incision and sent to pathology. Rarely, the symptoms of a blood-stained nipple discharge may be due to ductal carcinoma *in situ* (DCIS).

Epithelial hyperplasia

Epithelial hyperplasia (previously termed epitheliosis or papillomatosis) is usually an incidental finding on biopsy and

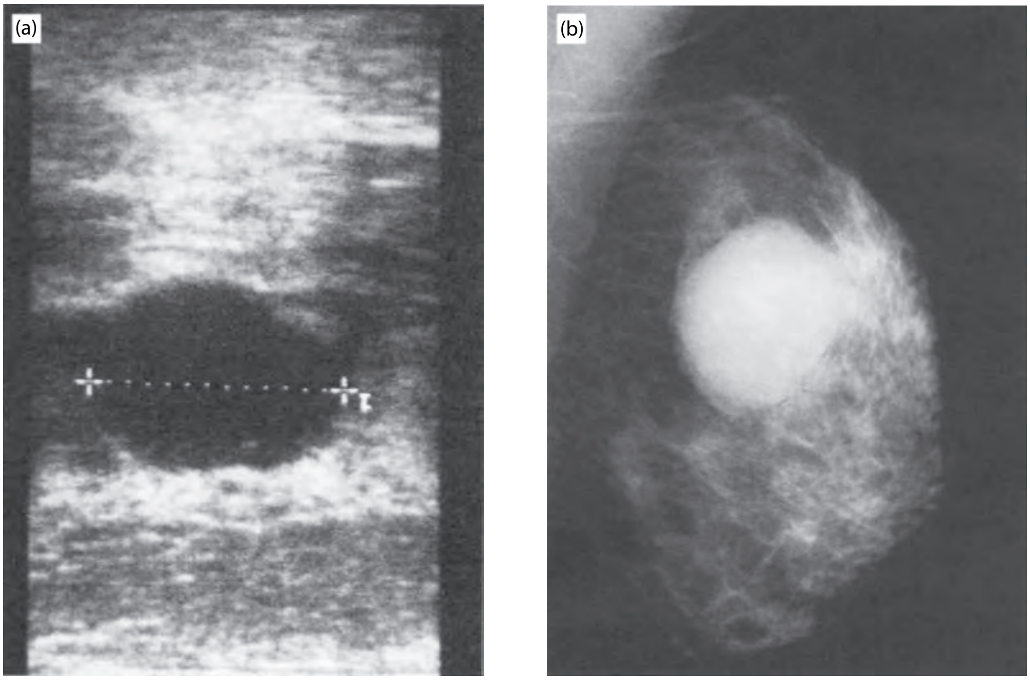


Figure 17.3 Cyst: radiological appearances using (a) ultrasound, showing an anechoic lesion with distal bright up; (b) mammogram of a breast cyst showing a large well-defined, oval mass.

can be graded as mild, moderate or florid. Cellular atypia (atypical hyperplasia) increases the risk of breast cancer two- to fourfold.

Sclerosing lesions

Usually detected on mammography as areas of architectural distortion, radial scars (which if >2 cm are also known as complex sclerosing lesions) require either surgical biopsy or vacuum biopsy excision to exclude cancer, which is present in about 15% of cases.

Inflammatory/infectious processes

Abscess (Figure 17.4)

Abscess formation may occur in the breast tissues or in the skin overlying the breast (in an epidermoid cyst) or in the inframammary fold. The typical symptoms of pain, redness and swelling are present.

A breast abscess is commonly lactational, occurring in a young breastfeeding woman, and usually staphylococcal in origin. Non-lactational abscesses tend to occur in middle-aged women who smoke and are secondary to streptococci and anaerobic bacteria. Periareolar infection, occurring in women in their thirties, results from active periductal inflammation (periductal mastitis). Periareolar abscess formation may proceed to a fistula (see below). Peripheral non-lactating breast abscess may be associated with diabetes, steroid therapy or rheumatoid arthritis. There may be multiple peripheral abscesses requiring drainage and antibiotics. At an early stage, abscess formation may be aborted by the use of appropriate antibiotics (such as amoxicillin, safe in breastfeeding mothers, with its antianaerobic and anti-staphylococcal action). Once an abscess is established, antibiotics are required for the surrounding cellulitis plus drainage either under local anaesthetic using a 19 G needle or by formal incision and drainage under local or general anaesthesia. Repeat drainage under ultrasound control helps prevent the need for surgical drainage in most cases. Inserting a drain, as described in older surgical texts, is *not* required. A breast abscess may require more than one drainage procedure (particularly if needle aspiration is used) and may be complicated by fistula formation. Tuberculosis and actinomycosis are rarer but recognized causes of breast abscess.

Inflammatory breast cancer should be excluded in patients with a solid inflammatory lesion or an apparent abscess that does not respond to appropriate treatment with ultrasound-guided drainage.

Fistula (Figure 17.5)

A fistula between the epithelium of a breast duct and the skin may follow on as a consequence of an abscess with the discharge site/fistula usually opening at the areolar margin. Excision of the fistula through a circumareolar incision including the duct up to the back of the nipple under antibiotic cover is the treatment of choice, rather than laying open the length of the fistula with the resultant scarring and deformity.

Duct ectasia (periductal mastitis)

These represent a spectrum of inflammatory process in which the subareolar ectatic, enlarged, ducts are surrounded by a mild inflammatory infiltrate commonly found in smokers. The ectatic

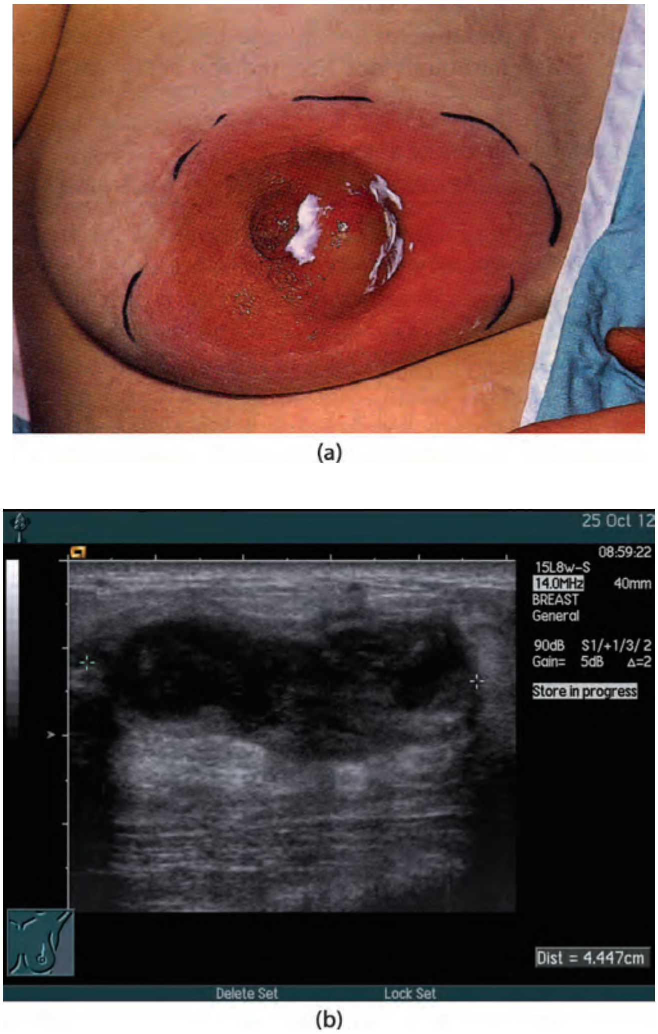


Figure 17.4 (a) Breast abscess in a breastfeeding mother demonstrating erythema marked to determine whether it is a spreading, pointing abscess adjacent to the areola with local anaesthetic cream in place prior to needle aspiration. (b) Ultrasound image of an abscess showing an ill-defined fluid collection with surrounding oedema.

duct may result in a retracted slit-shaped nipple (in contrast to the nipple retraction seen with an underlying cancer). Creamy discharge from multiple bilateral ducts is a feature of duct ectasia and is distinct from the single blood-stained discharge of a duct papilloma/intraduct carcinoma. Duct ectasia often shows characteristic 'lead pipe' and 'broken needle' calcifications on mammography.

Fat necrosis

Fat necrosis in the breast usually appears in postmenopausal women following trauma to the breast (such as a seat belt injury) as a localized inflammatory response. Fat necrosis may be suggested from the history but should be diagnosed using triple assessment including core biopsy as a small cancer may clinically and radiologically mimic fat necrosis. Sometimes fat necrosis, which often has a characteristic ultrasound appearance, can be confirmed by aspiration of associated oil cysts. Treatment is by reassurance or, if suspicion remains or the patient requests, excisional biopsy.

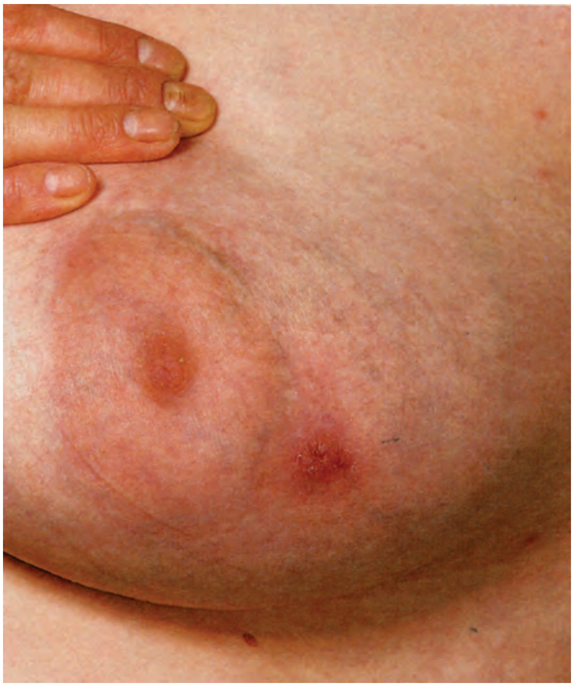


Figure 17.5 Fistula as a consequence of a breast abscess; note the classical position at the areolar margin.

Hormonal changes

Nipple discharge

Single or multiple duct discharge producing a clear, creamy, green or black fluid is a common physiological finding but also occurs from ectatic ducts. Physiological nipple discharge from multiple ducts, usually bilateral, is not uncommon in premenopausal women. Following triple assessment, including mammography to detect underlying DCIS, a careful drug history (particularly for psychotropic drugs, which may cause hyperprolactinaemia) and measurement of serum prolactin to exclude a (rare) pituitary prolactin-secreting tumour, surgical intervention (duct excision or disconnection) is only required if blood can be detected on stick testing raising the suspicion of an underlying papilloma or rarely a carcinoma, or if there is a profuse discharge causing patient distress.

Breast pain (mastalgia)

Breast discomfort or pain at some time is reported by two-thirds of women; mastalgia can be defined as breast pain of sufficient severity for a woman to seek medical advice. While breast pain is very rarely a symptom of breast cancer, appropriate reassurance is required.

Mastalgia may be unilateral, bilateral, unifocal or multifocal, cyclical (worse before the menses) or acyclical. It is important to exclude concomitant pathology. Cyclical mastalgia is worst in the days before the menses, usually affects the outer half of the breast, occurs in women often in their thirties and may persist for months or years. Cyclical mastalgia ceases with the menopause. Acyclical mastalgia may be continuous or have a random pattern and is more common in women in their forties. Hormone replacement therapy or a change in oral contraceptive pill may exacerbate mastalgia, at least initially. Following triple assessment of any focal abnormalities to exclude malignancy, conservative measures include reassurance and explanation, simple analgesia (e.g. paracetamol), dietary modification (avoiding caffeine in tea, coffee and soft drinks; avoiding fatty foods), wearing a correctly fitting supportive bra especially at night, stopping smoking and regular exercise. Keeping a breast pain chart (Figure 17.6) can be helpful for the patient to record the timing and severity of the pain and for the clinician to establish any pattern of pain.

Although mastalgia is self-limiting, it may persist for months or years. For cyclical mastalgia, medication such as gammalinoleic acid (oil of evening primrose) given for a minimum of 3 months' duration in an adequate dose (80 mg t.i.d.) shows a response similar to placebo in 40% of patients. If the pain is persistent and severe, medical therapy such as 100–200 mg danazol or bromocriptine have similar response rates, but one-third of women experience side effects. Tamoxifen 20 mg/day (although not licensed for breast pain) or ovarian suppression (by a luteinizing hormone-releasing hormone analogue) in premenopausal women is increasingly used for refractory mastalgia; psychological referral for the management of the pain may be required. Rarely, mastectomy is performed to remove the painful breast, but should be considered only when chest wall disease has been excluded, other measures have been exhausted and following patient counselling, as surgery tends to make mastalgia worse not better.

Acyclical mastalgia may respond to the conservative measures used for cyclical pain. A persistent localized painful area may respond empirically to local anaesthesia/steroid injection.

Chest wall pain, musculoskeletal pain (e.g. Bornholm myalgia), costochondritis (particularly of the second rib/cartilage joint – Tietze syndrome), thrombophlebitis of the chest wall veins (Mondor's condition; Figure 17.7), intra-abdominal (gall stones) or intrathoracic (cardiac, pleural) pain may all present as 'breast pain' in clinical practice. Exclusion of significant breast pathology, reassurance and non-steroidal anti-inflammatory drugs can provide symptomatic relief.

Other conditions

An accessory nipple, accessory breast tissue (which occurs along the axillary–groin milk lines) or, conversely, breast hypoplasia (which may be associated with pectoral/upper limb abnormalities – Poland syndrome) are developmental abnormalities of variations in the site and the volume of breast tissue. Surgical treatment is only required when swollen accessory breast tissue is symptomatic. Rarely, hamartoma, angioma or neurofibroma may occur in the breast.

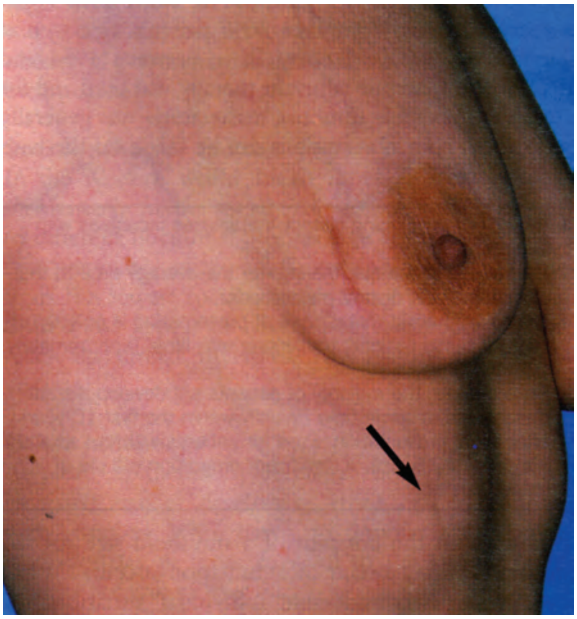


Figure 17.7 Mondor's disease (thrombophlebitis) inferior to the breast on the chest wall and a benign breast biopsy scar in the skin lines.

Lipoma

Lipomata are not uncommon in the breast and can mimic other breast lumps including carcinoma. They usually have a characteristic appearance on ultrasound and infrequently require biopsy for confirmation.

Epidermoid cyst

This lies within the skin of the breast, usually adjacent to the sternum or an inframammary fold. The classic punctum is often visible and the cyst may become inflamed/infected and discharge cheesy material. This lesion may be left *in situ* if not complicated by recurrent infection.

Each day mark whether you have bad pain, no pain or just a little pain, using the symbols shown.

◩ Mild Pain ■ Severe Pain ☐ No Pain

Please put a **P** in the space provided on the day your period starts

Month ____________________________

Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Pain?																															
Period start?																															

Figure 17.6 Breast pain chart for a patient to complete indicating the severity of pain and timing in relation to menses; the first month's chart (usually completed for 3 months).

Surgery for benign breast conditions

Many benign breast conditions are self-limiting and following triple assessment the patient may require simple reassurance, conservative measures or a course of medication. Follow-up of patients with benign breast conditions is not usually required in the hospital setting. A benign lump may be excised (under local or general anaesthetic) at the request of the patient. Surgical scars should be placed circumferentially, in the line of skin tension (Figure 17.7) or at the areolar margin, at the peripheral margins of the breast or in the skin lines of the axilla.

Gynaecomastia (Figure 17.8)

Gynaecomastia, the growth of breast tissue in males, is a benign, reversible condition affecting 3% of the male population at any one time. Teenage boys undergoing puberty (pubertal gynaecomastia) and elderly men (senescent gynaecomastia) are most commonly affected and between them account for half of the patients with gynaecomastia. Gynaecomastia secondary to drug therapy, liver disease, primary hypogonadism (e.g. Klinefelter syndrome), testicular tumours, hyperthyroidism or renal disease should be excluded. Withdrawal of the causative drug (cimetidine, digoxin, spironolactone, anabolic steroids, oestrogen) should result in symptom resolution.

Clinical examination, revealing a usually tender, palpable lump deep to the areolar and often, surprisingly, unilateral, may be supplemented by mammography, ultrasound and core biopsy if male breast cancer is suspected. In young men, the testes should be examined to exclude testicular tumours and human chorionic gonadotrophin and oestrogen levels measured. Surgical excision (through a circumareolar incision), danazol or tamoxifen have all been used for the treatment of gynaecomastia.

Female breast cancer

Breast cancer is the most common cancer in women and is second only to lung cancer in cancer-related deaths. Fortunately, mortality rates are declining in breast cancer – a change attributed to an ongoing multidisciplinary approach to the management of the disease. This approach includes the use of diagnostic imaging for early detection in appropriate patients. Upon the pathological or histological diagnosis of breast cancer, most patients undergo surgical resection and further pathological evaluation of breast and lymph node tissue followed by additional therapies such as radiation therapy and/or systemic therapy as indicated to reduce the risk of disease recurrence. Such an approach has been enhanced by the ability to identify high-risk patients who may especially benefit from increased clinical and radiographic surveillance to detect early disease or breast cancer prevention strategies which involve prophylactic surgery or drug administration.

Histology of breast cancer

Breast cancers originate from the epithelium of the terminal duct lobular unit. Those encompassed within the basement membrane are classified as *in situ,* and have characteristic patterns, as do invasive cancers that disseminate beyond the basement membrane.

Invasive cancers are now classified as no special type (NST; or not otherwise specified) (Figure 17.9) or special types: lobular, tubular, medullary, mucoid and papillary, which may have characteristic histopathological features and some of which have a better prognosis. Prognosis is related to lymph node involvement, tumour size and tumour grade: I, II, III based on gland formation, nuclear pleomorphism, and frequency of mitoses originally described by Scarff, Bloom and Richardson. These prognostic factors may be scored and combined to give a prognostic score. For invasive cancer, some 80% are NST, 10% lobular and the remainder (10%) tubular, medullary, mucoid, papillary or other even rarer types. The frequency of special type tumours is higher in women with screen-detected cancer. Lymphatic or vascular invasion on histological assessment is also a marker for both local and systemic recurrence. Patients with perineural invasion or extensive (>25%) *in situ* cancer within the tumour mass are more likely to develop local recurrence.

Paget's disease is an eczematous infiltration of large breast cancer cells into the nipple/areolar skin that occurs in approximately 2% of presenting cancers. It is symptomatic of an underlying high-grade ductal carcinoma *in situ* (DCIS) lesion which may have an invasive component. Diagnosis is from nipple or punch biopsy after triple assessment, which may reveal an underlying cancer.

Metastatic spread from breast carcinoma is via lymphatic channels to regional lymph nodes (Figure 17.10), predominantly in the axilla. Even in the absence of detectable lymph node metastasis,

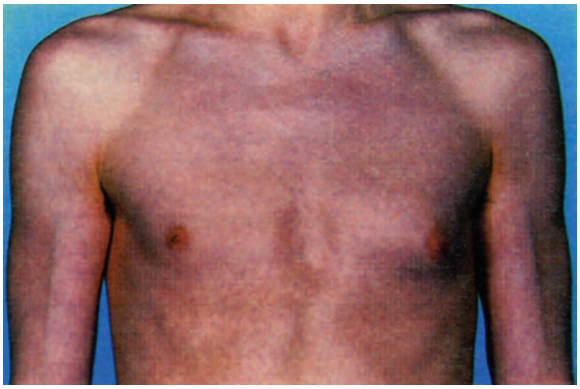


Figure 17.8 Unilateral (left) gynaecomastia in a teenage boy.

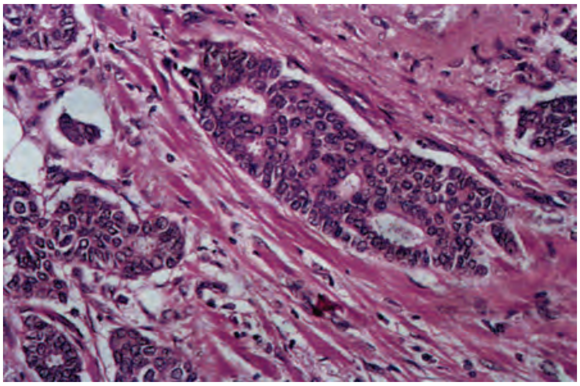


Figure 17.9 Haematoxylin and eosin-stained section of invasive breast cancer.

lymphatic or vascular invasion in the breast tissue is suggestive of the propensity of the cancer to spread to distant metastatic sites.

Synoptic reporting is proving increasingly acceptable to standardize and optimize communication, with the format determined locally for macroscopic and microscopic details. An illustrative example is provided in Box 17.5.

Carcinoma *in situ*

A spectrum of preinvasive neoplastic change in the breast includes DCIS (15% of symptomatic, 20% of screen-detected 'cancers'), lobular carcinoma *in situ* (LCIS; <1% symptomatic and 1% of screen-detected lesions) and hyperplastic appearances (ductal or lobular atypia), which can cause diagnostic difficulties.

DCIS (Figure 17.11) covers a heterogeneous group of lesions and is classified by histological pattern, grade (high, intermediate or low) and the presence of necrosis. However, no internationally agreed classification system exists. High-quality mammography is required to determine the extent of disease (but may still underestimate the volume of breast involved), although DCIS can be present without mammographic signs. The majority of cases are now detected in screening programmes, although the natural history of the disease is not fully understood.

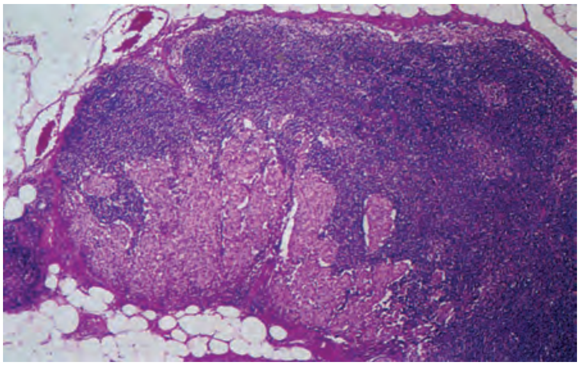


Figure 17.10 Haematoxylin and eosin-stained section of an axillary lymph node showing extensive node involvement with invasive breast cancer.

BOX 17.5 Synoptic pathology report

- Invasive cancer: invasive carcinoma, no special type
- Grade: III (glands 3, nuclei 3, mitosis 2)
- Size: 27 mm
- *In situ* cancer: present – comedo type
- Grade: high
- Extent: extensive, within main lesion
- Calcification: yes; associated with *in situ* cancer
- Margins: nearest is medial at 5 mm, all others >10 mm
- Multifocal: no
- Vascular/lymphatic invasion: yes, at the edge of the cancer in two blocks
- Lymph nodes: number confirmed 5; number positive 1 (microscopic tumour deposit)
- Oestrogen receptor (specify scoring system): Allred 6/8, HER2: positive 3^+ FISH amplified
- Summary: right breast wide local excision – invasive carcinoma grade 3 with one lymph node metastasis

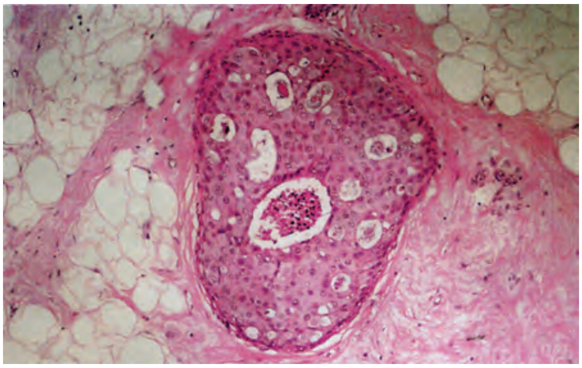


Figure 17.11 Haematoxylin and eosin-stained section demonstrating ductal carcinoma *in situ* (of cribriform type).

DCIS, characterized by ducts expanded by large irregular cells with large irregular nuclei, may be classified in a variety of ways: comedo DCIS has high-grade cytology, extensive necrosis and branched calcifications; non-comedo DCIS (cribriform, solid or micropapillary) has low-grade cytology, lacks necrosis and calcification is inconsistent. As one might expect, there are patients with intermediate histology. DCIS is increasing in the frequency of detection, largely because of screening techniques. The risk of developing invasive cancer at the site of low-grade DCIS is about 10% over a 10 year period. The natural history of high-grade DCIS is more aggressive with a median time to relapse of 76 months. The average age of diagnosis is 55 and most women are postmenopausal. The biology remains poorly understood (e.g. 60% of DCIS is HER2 protein expressing (see below), but, unlike invasive cancer, this is not a poor prognostic sign).

Atypical ductal hyperplasia

Atypical ductal hyperplasia (ADH) forms part of a spectrum with DCIS – membrane-bound spaces of 2–3 mm with cellular atypia. ADH holds a fourfold risk of developing breast cancer, which is additive with any family history of breast cancer.

Lobular carcinoma *in situ*

LCIS, usually an incidental histological finding, is an expansion of the breast lobule by smaller, regular cells with regular row/oval nuclei. The risk of developing invasive breast cancer is about 25% (and contralateral breast cancer about 10%) at 20 years; since most women (70%) with LCIS are premenopausal (mean age 45 years) this may be significant.

Lobular neoplasia, identified on core biopsy, carries an upgrade rate at surgical excision to DCIS or invasive carcinoma of around 15%. Such lesions are therefore usually excised.

Other malignancies in the breast

Primary lymphoma affecting nodes within the breast or in the axilla is perhaps the most common non-breast malignancy and is treated, following staging, with or without local excision/tumour biopsy, by conventional antilymphoma regimens. Sarcoma is treated by excision with the widest possible margin at mastectomy followed by postoperative radiotherapy.

While primary breast cancer is by far the most common breast malignancy, secondary spread from other organs such as bronchogenic carcinoma, melanoma or from the gynaecological organs is recognized. Usually, the cellular morphology of the lesion gives clues to its origin (which may be suggested from a past history of, for example, leiomyosarcoma of the uterus), and the use of special stains suggesting the lesion is not breast cancer [e.g. oestrogen receptor (ER) negative] or positive molecular markers (e.g. for lung cancer) give clues to the primary.

Molecular characterization of breast cancer

There appears to be a wide range of molecular abnormalities in breast cancer which probably interact with the cellular environment and play a role in the aetiology, development, invasion and metastasis of breast cancer. Using whole genome scanning techniques, such as comparative genomic hybridization or genetic profiling (with microarray or chip-based technology for gene expression or gene deletion), it is clear that individual cancers vary markedly from each other. In clinical practice, few genes or their protein products are of genuine widespread clinical utility but include the ER gene protein product and a cell surface receptor HER2. Some provide information when the importance in clinical decision-making (PgR, pS2/PNR2, *p53*) remains unclear, whereas others (Ki67 or PCNA as a proliferation marker, percentage of cells in S phase, DNA content of tumour) relate to the biological aggressiveness of the cancer.

Oestrogen receptor

ER expression, historically measured by radioimmunoassay or the dextran-coated charcoal method on cell lysates, is now assessed by immunohistochemistry on formalin-fixed paraffin-embedded tissue sections (Figure 17.12), for which there are a variety of scoring systems. Two forms of ER exist: ER-α (detected and measured clinically in breast cancer samples) and ER-β (of uncertain clinical significance in breast cancer); tumours with moderate or high ER expression are much more likely to respond to endocrine therapy than those with no or low ER expression. In addition, ER expression correlates with survival, at least in the first 5 years following diagnosis. However, only two-thirds of patients with ER expression appear to have intact, functioning downstream effector pathways, so some centres also measure progesterone receptor (PgR) and/or the insulin-like growth factor peptide pS2 in immunohistochemical sections. Expression of PgR and/or pS2 along with ER suggests that the cancer is more likely to respond to endocrine therapy.

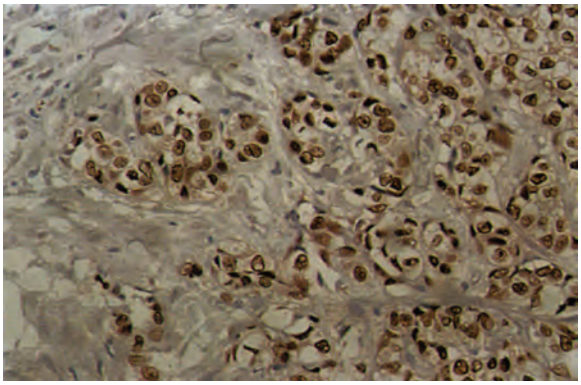


Figure 17.12 Oestrogen receptor staining of breast cancer.

Human epithelial growth factor receptors

HER2 (erbB2, neu) is a transmembrane growth factor receptor, expressed in 15% of invasive breast cancers (and, for reasons which are unclear, up to 60% of DCIS), and is associated with poor prognosis in invasive breast cancer. It is one of a family of growth factor receptors [including epidermal growth factor receptor (EGFR), erbB3, erbB4], but unlike most other molecular markers now forms a target for ‘biological therapy’. For patients in whom amplified HER2 can be detected (i.e. multiple copies for the gene best detected by fluorescence *in situ* hybridization) and is overexpressed (on immunohistochemistry), humanized mouse antibody [trastuzumab (Herceptin)] was the first new ‘biological’ therapy to enter clinical use. What is perhaps of greatest clinical significance is that this therapy may be most effective in HER2-positive and node-positive patients with invasive breast cancer whose disease is relatively insensitive to chemotherapy or endocrine therapy. A number of targeted therapies to HER2 extracellular or intracellular domains or HER2 and EGFR (targeting EGFR–HER2 dimers which drive HER2 downstream effects) have subsequently emerged.

p53

p53 has acquired the nickname ‘guardian of the genome’ owing to its essential role in the cellular response to injury with subsequent cell cycle arrest or apoptosis. In breast cancer, mutant *p53* has been associated with poor prognosis and resistance to treatment by chemotherapy. Thus, detection of *p53* mutation may be important in clinical therapeutic decision-making and is the subject of clinical trials. Unfortunately, immunohistochemistry for *p53* in breast cancer is less reliable than other methods of detecting *p53* mutation and this has held back the application of *p53* research in the clinical arena.

Triple negative and basal phenotype breast cancer

Cancers which are negative for ER, PgR and HER2 are termed triple negative. Most triple negative cancers express basal cytokeratins such as CK5/6 and CK14. Triple negative and basal breast cancers have a poor prognosis even when detected at small sizes.

Triple negative and basal breast cancers are more frequently found in *BRCA1* carriers and women of African Caribbean or African American ethnicity. Basal breast cancers present at a younger age than luminal (ER positive) breast cancers. Targeted therapy is currently not available but the use of poly ADP ribose polymerase (PARP) inhibitors is being investigated in this patient group.

Other genes

Several other genes [multiple drug resistance (*mdr*) glycoprotein], oncogenes (c-*myc*), apoptotic pathway genes (*bcl2*, *bax*), tumour suppressor genes (*Rb*), growth factors (transforming growth factor α and β families, insulin-like growth factors), growth factor receptors (EGFR), proteases (cathepsin D) and cell cycle-related genes (*p21*, cyclin D) have been among the many studied in breast cancer. Each has its enthusiasts, but to date they have added little to clinical practice.

Gene signatures

Combining a series of genes into a predictive and/or prognostic profile has resulted in the generation of some dozen or so molecular tests. Many are based on the RNA extracted from frozen or formalin-fixed paraffin-embedded tissues whereas others have used immunohistochemistry predictors. While these technologies remain the subject of clinical trials, they aim to predict which patients may not need chemotherapy (the MINDACT trial based on frozen material and the MammaPrint diagnostic platform) or which ER-positive patients may benefit from chemotherapy (the TAILORx trial using the Oncotype DX test and formalin-fixed, paraffin embedded material). The clinical use of these technologies has not been well established worldwide and their cost together with the need for further evidence of their utility has held back the introduction of such tests into everyday clinical practice.

Epidemiology of breast cancer

It is estimated that the world burden of breast cancer is some 1 million women newly diagnosed each year. The extent of the breast cancer problem can be illustrated based on the data from the UK. Breast cancer is the most common incident cancer among women, with some 48 000 new registrations, and accounts for 25% of the female cancer burden excluding non-melanoma skin cancer. For women, the lifetime risk of developing breast cancer is 11%; it is the commonest cause of death in women aged 40–50 years.

Breast cancer is the second most common cause of death from cancer in women (after lung cancer). Survival has improved over the past 20 years in, for example, Scotland, with 56% 5 year relative survival having been reported for those diagnosed between 1968 and 1972, compared with 70% for those diagnosed between 1988 and 1992, rising to 90% for women diagnosed in 2004. However, even allowing for pathological stage, there is evidence of variation between different patient groups.

The rising incidence of breast cancer may be due to the increasingly aged population together with earlier detection (including DCIS, which may never develop into invasive cancer) through screening programmes in addition to the factors listed below. This increasing incidence may be offset by the improved survival secondary to the effect of early detection of small cancers prior to metastasis and the use of adjuvant endocrine therapy, radiotherapy, chemotherapy and biologically targeted therapy.

Aetiological factors in breast cancer

Although the aetiology of breast cancer is multifactorial and may differ between individual women and racial groups, it is clear that breast cancer has a genetic basis upon which internal hormonal influences, diet and external environmental factors act.

Endocrine influences on breast epithelium result in stimulation, then involution, of breast tissue. The number of cycles of stimulation/involution together with the stimulatory effect of exogenous steroid hormones (oestrogen, progesterone) may additively increase the risk of breast cancer.

Age

The risk of breast cancer increases with age, although there is a flattening of the age–incidence curve after the menopause. The median age for breast cancer in Western countries is 60 years.

Hormonal influences

These include:

- Age at menarche (younger age increases the risk).
- Age at menopause (menopause at age 55 doubles the risk compared with menopause at age 45).
- Number of pregnancies (higher number is protective).
- Age at first pregnancy (if over 30, the risk is double that at <20 years at first pregnancy).
- Breastfeeding history (breastfeeding is protective).
- Oral contraceptive use (high oestrogen pills increase risk, whereas low oestrogen/progesterone pills are safer; use before age of 20 increases the risk; note that oral contraceptive use decreases the risk of ovarian cancer). Risk is elevated during and for 10 years after use.
- Hormone replacement therapy (HRT). A recent analysis of pooled published data on HRT and the risk of breast cancer has shown that the relative risk for each year of use is 1.023 and for 5 years or more of use the relative risk is 1.350. For women under the age of 50 this risk is equivalent to having normal menses and therefore women under 50 are at no increased risk. Five years after cessation of HRT, the risk returns to normal levels. A baseline mammogram prior to commencement of HRT is not recommended. Some studies have suggested that cancers diagnosed in women on HRT may be at an earlier stage but the Million Women Study has shown increased breast cancer mortality in current users of HRT. Breast cancer risk from HRT should be balanced against the potential benefits of HRT, including control of menopausal symptoms and osteoporosis.
- Oophorectomy in premenopausal women greatly reduces the risk of breast cancer, the benefit being greatest when oophorectomy is conducted at a young age, presumably by reducing the cyclical oestrogenic stimulation of breast tissue.

Family history

A family history of breast, ovarian, colon, prostate or other cancers is associated with increased risk of breast cancer.

When a family member has had breast or ovarian cancer, their age at onset should be recorded as well as whether the cancer was unilateral/bilateral, the menopausal status of the family member and the relationship to the individual consulting – first-degree relative (mother, sister or daughter); second-degree relative; or maternal or paternal side. There are now guidelines (Box 17.4) for patients to attend risk assessment clinics for screening which, when conducted at centres with expertise and interest in the field, can detect cancers with similar success to breast screening programmes conducted in postmenopausal women.

Mammographic density

Women whose mammogram shows 75–100% dense tissue are at four to five times increased risk of breast cancer compared with women with <10% dense breast tissue. Decreasing breast density in middle life has recently been shown to be associated with a decreased risk of breast cancer.

Diet

Dietary influences on breast cancer risk are most startling when noting the increased risk of breast cancer in Japanese women moving to the USA. Some evidence suggests that soya and phyto-oestrogens are protective, but eating saturated fat and red meat, drinking alcohol (for which there is a dose–response relationship), obesity in postmenopausal women and smoking all increase the risks of breast cancer.

Socioeconomic factors

In the Western world, breast cancer has an increased incidence in higher socioeconomic groups of women, possibly related to different hormonal profiles (age at menarche, age at first pregnancy, etc.) in different social classes. Women from the lowest socioeconomic groups present with later stage disease and appear to have a poorer outcome, despite receiving similar therapies. The reasons for this are unclear.

Other important influences

Atypical epithelial hyperplasia holds a fourfold increased risk of developing breast cancer compared with women who have no proliferative changes. Therapeutic radiation for lymphoma, given as thoracic irradiation in teenage or young women, significantly increases the risk of breast cancer. Smoking now appears to increase the risk of breast cancer.

Familial predisposition

About 5–10% of breast cancers in the West are attributed to a genetic predisposition – there are multiple genes of limited penetrance, inherited in an autosomal dominant fashion.

However, *BRCA1* (17q) and *BRCA2* (13q) account for many of the high penetrant larger families with multiple (four or more) breast and other cancers in close relatives. Within certain populations, the particular mutation identified (e.g. *BRCA2* deletion at position 999 in half of familial cancers in Iceland) demonstrates the genetic lines of inheritance. However, population screening for mutations in *BRCA1* and *BRCA2* is not, at present, realistic given the large size of these genes and the scatter of mutational sites. Inheritance of mutant *p53* (Li–Fraumeni syndrome) or *PTEN* (Cowden syndrome) is even less common. These four genes have comparatively high penetrance resulting in multiple cancer sites within each family and sometimes within each individual. The breast cancers usually occur before the age of 65 years and may be bilateral. Based on a detailed family history, risk estimates can be computed using established tables to help inform women of whether their risk is sufficient to warrant intervention. Other genes such as ataxia telangiectasia (*ATM*), *FGFR2* and other low penetrative genes probably account for many of the smaller families (two or three cases) in which *BRCA1* and *BRCA2* have been excluded.

Early detection and prevention of breast cancer

Based on the concept that detecting and treating precancerous lesions and small cancers (before they have metastasized to the regional nodes and/or further afield) could save lives, breast screening and, more recently, (chemo)prevention of breast cancer in high-risk groups have been implemented.

Breast screening

There is currently no evidence to support the view that clinical breast examination by a doctor, nurse or patient should be considered as a primary screening technique. However, since the majority of breast cancers are found by women themselves, women should be encouraged to become aware of the feel and shape of their breasts, so that they are familiar with what is normal for them and to report any change from normal to a medical practitioner.

National breast screening programmes, such as in the UK, performed by mammography, aim to reach >70% of the target population and require considerable organization and a built-in quality assurance programme. Digital two-view mammography (craniocaudal and oblique), with double reading of films performed every 2–3 years, should maximize the detection of small (<15 mm) breast cancers.

The UK National Health Service Breast Screening Programme (NHSBSP) aims to detect 7.3 cancers/1000 women screened of both the first (prevalent) attendence at age 50–52 and on average at subsequent incident rounds. Eighty percent of the cancers detected will be invasive, 20% non-invasive, of which nearly all are DCIS. More than 50% of the cancers detected should be <15mm in pathological size.

Some two-thirds of screen-detected abnormalities prove to be insignificant on further mammographic/ultrasound review. For the remainder, triple assessment is required, and, since two-thirds are impalpable, ultrasound- or mammographic-guided biopsy with core or vacuum (wide bore core) biopsy should establish the diagnosis. Mass lesions and distortions should initially be biopsied using core biopsy whereas vacuum biopsy is the biopsy method of choice for calcifications. Lesions of uncertain malignant potential (papillomas and radial scars) but without cellular atypia have a low upgrade rate and can be removed using vacuum excision, whereas lesions with cellular atypia have higher upgrade rates and should be removed at surgery. For invasive malignant lesions a preoperative diagnosis rate of 95% should be achieved. Preoperative diagnosis rates for DCIS are lower at around 85%.

Screen-detected cancers are, in general, smaller, better differentiated, more likely to be of special type and node negative than symptomatic cancers; these features agree with the concept of detecting 'early' cancers, when the chances of metastasis are less and mortality is lower. Thus, population screening by mammography can reduce mortality from breast cancer by up to 25% in those who attend.

In the UK, women in the 50–70 years age range are invited every 3 years for screening through the NHSBSP aiming to achieve a 29% reduction in mortality. Meta-analyses of the international screening mammography trials also show statistically significant mortality reduction of 18–29% in the 40–49 years age group. The efficacy of screening women in their forties in the UK has been evaluated in the UK Co-ordinating Committee on Cancer Research (UKCCCR) Age trial. This study showed a non-significant 17% mortality reduction. Thus, with evidence that screening can be effective in the 40–70 years age group, there is the potential for reducing the mortality of breast cancer by using population-based screening. Unfortunately, this approach cannot be adopted throughout the world, owing to constraints on resources or the pattern of clinical care in some countries.

Radiation risk and mammography

It is thought that ionizing radiation increases the risk of breast cancer development after a latent period of 10 years, that the risk is cumulative and that the risk is greatest for adolescent exposure. This is based on the increased incidence of breast cancer in women who have had therapeutic thoracic radiotherapy for lymphoma, and decreases with increasing age at exposure. In those aged over 50 years, the risk of cancer induction is, very approximately, 1:100 000 per single view mammographic examination. The average dose per examination (single view per breast) is approximately 1–2 mGy, the dose being dependent on breast thickness and exposure factors used.

Women at increased risk of breast cancer

It is estimated that some 5% of all cases of breast cancer are attributable to inheritance of a gene conferring a high lifetime risk of breast cancer. A characteristic of many cases of genetically determined breast cancer is early age of onset. Several genes predisposing to breast cancer have been identified including *BRCA1*, *BRCA2*, *ATM* and *p53*.

Some women with histologically identifiable lesions, e.g. LCIS or severe atypical hyperplasia, are at a higher relative risk (of about two- to fourfold) of breast cancer. The risk is further increased when there is a positive family history. Thus, women with LCIS or ADH should have annual or biennial mammography. DCIS is considered below.

Only those women judged to be at substantially increased risk (Box 17.4) should be considered at present for detailed genetic assessment and follow-up in specialist clinics. In families with four or more relatives affected with either breast or ovarian cancer in three generations and one living affected individual, direct gene testing might be appropriate.

Management of women at high risk of breast cancer

For some families, the risk decreases with age and this should be reflected in their management. Women at high risk of breast cancer have three options: increased frequency of screening, including the use of MRI; chemoprevention; or prophylactic surgery.

Increased frequency of screening

Regular clinical and mammographic examination has been offered to young women at high risk of breast cancer without controlled trials. The efficacy of screening of young, high-risk women is at present uncertain, but a recent review by the UK Familial Breast Cancer Group suggested targeted screening may be as efficacious in this age group as in the NHSBSP.

In the UK, it is suggested that mammography should start at the age of 35 years, or 5 years younger than the youngest affected family member, whichever is younger. For women at high risk of breast cancer:

- <40 years: biennial mammography and annual clinical examination
- 40–50 years: annual mammography and clinical examination
- >50 years: depending on the risk, either discharge to NHSBSP or continue more frequent screening.

In the UK, known gene carriers (or women at equivalent risk) who do not opt for prophylactic surgery are eligible for annual MRI screening. MRI screening in this patient group has been shown to have superior performance in detecting small node-negative breast cancers than ultrasound or mammography. Women with *BRCA1* mutations have a high frequency of basal phenotype tumours. Basal cancers, unlike other types of breast cancer, may not have improved survival if detected early, so the use of MRI screening in *BRCA1* carriers remains unproven.

Prophylactic surgery

Some women at very high risk (>35% lifetime risk) of breast cancer may wish to consider prophylactic bilateral mastectomy, usually with reconstruction. If subcutaneous mastectomy is considered, the risk from residual ductal tissue deep to the nipple or in the tail of the breast must be recognized as any familial genetic defect would theoretically be carried by all remaining cells. The consequent reduction in the incidence of breast cancer by 90% in such high-risk women is considered worthwhile.

Management of women with autosomal dominant inheritance who develop breast cancer

Causal mutations can be identified in some 20% of high-risk breast cancer families. Genetic testing must involve formal pre- and post-testing counselling in person by a clinical geneticist. Predictive tests should only be undertaken in diagnostic laboratories with appropriate quality control.

Women who carry a mutant *BRCA1* or *BRCA2* gene and develop early breast cancer have a risk of developing cancer in the contralateral breast of up to 64%. In such patients presenting with good prognosis tumours, bilateral mastectomy should be discussed with the patient by her surgeon.

The risk of ovarian cancer is dependent on the family history and gene involved. Both high-risk (63% lifetime risk at 70 years) and low-risk (<40%) families exist. Preliminary data suggest that prophylactic salpingo-oophorectomy does reduce the risk of ovarian cancer.

Assessment, investigation and staging of breast cancer

Methods of assessing breast conditions, including breast cancer, are based on triple assessment: clinical history and examination, imaging and pathological assessment.

Imaging of symptomatic breast disease

In patients with symptomatic disease two-view mammography should be performed: craniocaudal and oblique views (Figure 17.13) supplemented by compression/paddle or magnification views. The radiological classification of breast appearances is listed in Table 17.1. Mammography is not recommended under the age of 40 years unless there is a strong clinical suspicion of carcinoma. Ultrasound examination may provide additional information to mammography (Figure 17.3), and can be useful to visualize focal breast disease in women under 40 years in whom the breast is radiologically dense and lesions may be more difficult to detect. Thus, mammography alone may not exclude

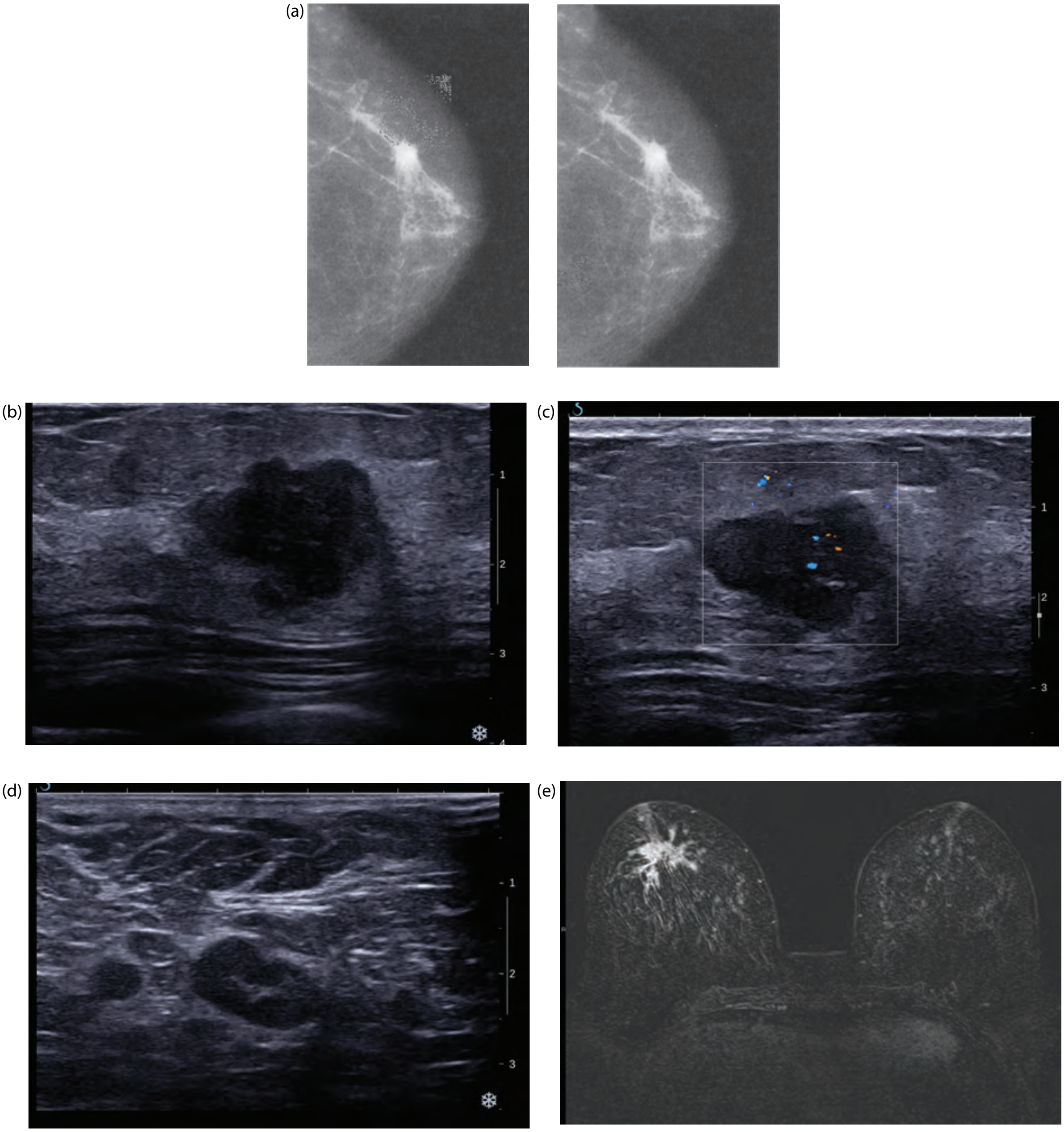


Figure 17.13 Images of breast cancer: (a) mammograms of a screen-detected breast cancer showing a spiculated mass. (b) Ultrasound image showing an ill-defined, inhomogeneous, irregular mass. (c) Colour Doppler image of a breast cancer showing central neovascularity. (d) Ultrasound of the axilla showing a node with a thickened cortex due to axillary metastasis. (e) MRI of invasive cancer.

Table 17.1 Classification of radiological appearances

Mammograms		Breast type		Ultrasound	
R1	Normal	N1	Good visualization; no masses/calcification/deformities	U1	Normal/diffuse benign
R2	Benign	P1		U2	Single cyst
R3	Indeterminate	P2		U3	Solid benign
R4	Probably malignant	PDY	Dense; cannot rule out masses	U4	Suspicious of malignancy
R5	Malignant	DY		U5	Malignant
				U6	Multiple cysts

breast cancer, particularly in younger women. Ultrasound is superior to mammography in characterizing breast masses. It is the primary imaging tool in women who are young, pregnant, lactating or have suspected infection. Ultrasound is also useful in the preoperative assessment of the axilla in breast cancer patients. Ultrasound has improved in recent years with additional techniques such as power Doppler and elastography.

Stereotactic (mammographic) or ultrasound-guided core biopsy should be performed on impalpable lesions that are suspicious or equivocal on radiological review. This may provide sufficient information to allow a benign lesion to be left *in situ*, but, if clinical doubt persists or the patient requests, needle localization biopsy of the impalpable lesion with specimen radiographs is necessary to allow histological examination of the appropriate portion of the biopsy specimen. Palpable lesions are also best biopsied under ultrasound control.

Other imaging modalities

MRI has a number of well-defined indications:

- sizing lobular cancers prior to wide local excision or mastectomy
- sizing breast cancer when conventional assessments are difficult or conflicting
- indentifying the primary tumour in women with metastatic adenocarcinoma in the axillary nodes and normal mammography, breast ultrasound and clinical breast examination
- identifying intracapsular rupture of silicone implants
- assessment of response in women receiving neoadjuvant chemotherapy
- excluding invasive cancer in women with mammographically detected lesions not amenable to image-guided biopsy.

Positron emission tomography/CT, breast CT and breast-specific gamma imaging are promising imaging modalities but their place in routine clinical practice is yet to be defined.

Histological diagnosis of breast cancer

Core biopsy of a focal breast abnormality using a 14 G needle with a 22 mm throw is the best method of obtaining a histological diagnosis of most breast lesions. The reported samples are classified as B_1 (normal tissue), B_2 (benign), B_3 (lesion of uncertain malignant potential), B_4 (suspicious of cancer), B_{5a} (DCIS) or B_{5b} (invasive cancer).

B_3 lesions are subdivided on basis of epithelial atypia. Lesions without atypia have low upgrade rates at surgery so vacuum removal is an option. Lesions with atypia have higher upgrade rates and require surgical excision. Lesions yielding a B_4 result are often repeated to try to make a definitive diagnosis of malignancy. Immunohistochemistry to give ER, PgR and HER2 status can be performed on the core samples giving a B_{5b} result. Provisional grade and typing of invasive cancers and DCIS can be performed on core biopsy samples. Preoperative diagnosis rates for DCIS using core biopsy are not high, so many units now use 11 G vacuum biopsy devices, which are associated with higher diagnosis rates of DCIS and lower upgrade rates of ADH to DCIS and DCIS to invasion than core biopsy.

Staging of breast cancer

Clinical staging of breast cancer

Patients are staged clinically according to the Union for International Cancer Control TNM classification (Table 17.2). Although clinical assessment of tumour size is not completely accurate, and ultrasound, mammogram or MRI can be more accurate, there is reasonable correlation between clinical tumour size and final pathological assessment. All patients with a breast mass should have a careful clinical examination and the mass may be measured with callipers. The presence or absence of any signs of local advancement (inflammation, peau d'orange, ulceration, satellite nodules, direct chest wall involvement, fixed axillary nodes) should be noted and recorded. Clinical examination is an unreliable guide as to whether there is involvement of the axillary nodes by metastatic breast cancer. Ultrasound of the axilla and biopsy of abnormal nodes is able to give a preoperative diagnosis of nodal involvement in about 40% of cases.

In early, operable breast cancer (T1–2, N0–1), there is no current evidence to support routine screening for metastatic disease in asymptomatic women. Such patients should normally only have investigations to confirm general fitness, which may include a chest radiograph, ECG, full blood count, urea, electrolytes and liver function tests. Patients with symptoms suggestive of metastases require appropriate investigation. The incidence of asymptomatic metastases increases as the T (Figure 17.14) and N stage of the locoregional cancer increases. If it will affect treatment, patients with more advanced but operable disease (T3, N1–2) may require staging to exclude distant metastases. Screening for metastatic disease is best performed with chest, abdominal and pelvic CT scan. Such investigations aim to identify bone, visceral and soft-tissue metastases. If metastases are detected, the tumour markers CEA and CA15-3 may be elevated and can be useful for monitoring disease progression in certain patients.

Pathological staging of breast cancer

The pathology staging of breast cancer follows similar principles to clinical staging, but requires tissue to stage the tumour size

Table 17.2 Tumour (T)–node (N)–metastasis (M) staging

T Primary tumour		
TX		Primary tumour cannot be assessed
T0		No evidence of primary tumour
Tis		Carcinoma *in situ*: intraductal carcinoma, or lobular carcinoma *in situ*, or Paget's disease[a] of the nipple with no tumour
T1		Tumour 2 cm or less in greatest dimension
	T1mic	Microinvasion[b] 0.1 cm or less in greatest dimension
	T1a	More than 0.1 cm but not more than 0.5 cm in greatest dimension
	T1b	More than 0.5 cm but not more than 1 cm in greatest dimension
	T1c	More than 1 cm but not more than 2 cm in greatest dimension
T2		Tumour more than 2 cm but not more than 5 cm in greatest dimension
T3		Tumour more than 5 cm in greatest dimension
T4		Tumour of any size with direct extension to chest wall[c] or skin
	T4a	Extension to chest wall
	T4b	Oedema (including peau d'orange), or ulceration of the skin of the breast, or satellite skin nodules confined to the same breast
	T4c	Both 4a and 4b, above
	T4d	Inflammatory carcinoma[d]
N Regional lymph nodes		
NX		Regional lymph nodes cannot be assessed (e.g. previously removed)
N0		No regional lymph node metastasis
N1		Metastasis to movable ipsilateral node(s)
N2		Metastasis to ipsilateral axillary node(s) fixed to one another or to other structures
N3		Metastasis to ipsilateral internal mammary lymph node(s)
M Distant metastasis		
MX		Distant metastasis cannot be assessed
M0		No distant metastasis
M1		Distant metastasis

[a] Paget's disease associated with a tumour is classified according to the size of the tumour.

[b] Microinvasion is the extension of cancer cells beyond the basement membrane into the adjacent tissues with no focus more than 0.1 cm in greatest dimension. When there are multiple foci of microinvasion, the size of only the largest focus is used to classify the microinvasion. (Do not use the sum of all the individual foci.) The presence of multiple foci of microinvasion should be noted, as it is with multiple larger invasive carcinomas.

[c] Chest wall includes ribs, intercostal muscles and serratus anterior muscle but not pectoral muscle.

[d] Inflammatory carcinoma of the breast is characterized by diffuse, brawny induration of the skin with an erysipeloid edge, usually with no underlying mass. If the skin biopsy is negative and there is no localized measurable primary cancer, the T category is pTX when pathologically staging a clinical inflammatory carcinoma (T4d). Dimpling of the skin, nipple retraction or other skin changes, except those in T4b and T4d, may occur in T1, T2 or T3 without affecting the classification.

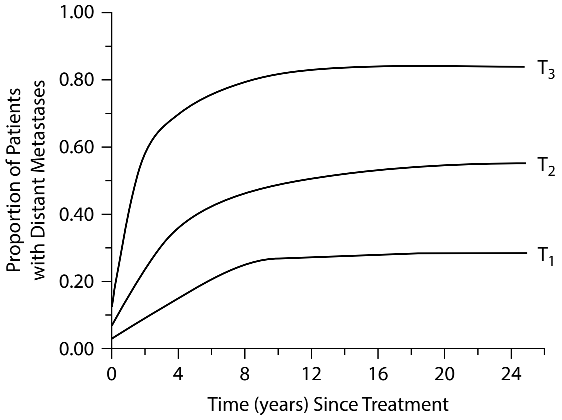


Figure 17.14 Tumour size in relation to proportion of patients with metastases following initial treatment for breast cancer.

and whether there is involvement of regional lymph nodes. The TNM stage is then adapted to pTpN and, when pathology material is available from a metastasis, pM. For lymph nodes the presence of micrometastases or isolated tumour cells may also be noted but are currently considered to be pN0.

Pathology

Prognostic factors

In the management of breast cancer, prognostic factors aid the selection of treatment for individual patients. Tumour size (measured histological larger size means worse prognosis), axillary node metastasis (Figure 17.15) (on pathology, not clinical, grounds, survival is directly correlated with the number of lymph nodes involved – women with >10 nodes do particularly badly), high histological grade, lymphatic and/or vascular invasion and the presence of distant metastasis are all important prognostic factors.

Prognostic indices

Combined with tumour grade, histological tumour size and pathological node status have been used to construct the Nottingham Prognostic Index (NPI), where NPI = (tumour size in cm × 0.2) + lymph node stage (1, no nodes; 2, one to three nodes positive; 3, four or more nodes involved) + grade (1, 2 or 3). Originally forming three prognostic groups (good, moderate, poor) the NPI is now used to place patients in one of five prognostic groups (Blamey, 1996) (Table 17.3).

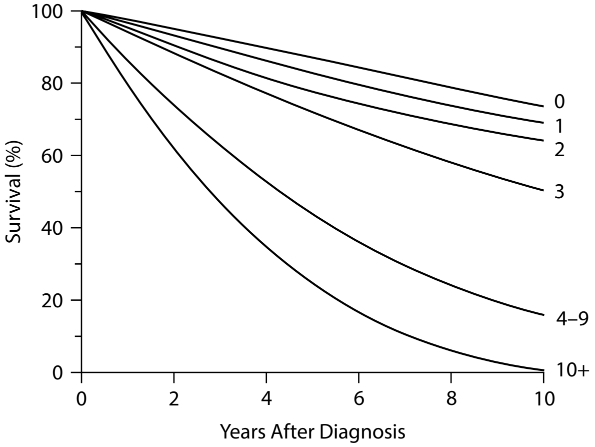


Figure 17.15 Axillary node involvement (number of nodes on pathology review) compared with survival following the diagnosis of breast cancer.

Table 17.3 Nottingham Prognostic Index (NPI) groups

Prognostic group	NPI	10 year survival (%)
Excellent	≤2.4	94
Good	≤3.4	83
Moderate I	≤4.4	70
Moderate II	<5.4	51
Poor	>5.4	19

Computational techniques have been used to develop the Adjuvant! online program, which incorporates additional information (such as the patient's age and ER status) and can estimate the likely benefit to a patient of receiving endocrine and/or chemotherapy. More recently, the PREDICT prognostication model has been used to predict breast cancer survival following surgery for breast cancer, based on UK cancer registry data, and can also take into account the mode of detection of the cancer.

High-risk subtypes of breast cancer

Locally advanced breast tumours

These are defined as primary tumours >5 cm in size (or fixed to skin or chest wall) or, less often, patients with bulky or fixed axillary lymph nodes. These patients would thus be staged as T3–4, N0–2, M0. As previously mentioned, there is a significant risk of these patients having detectable metastatic disease, and they should be formally staged by CT even if asymptomatic, to confirm the absence of overt metastases.

Some of these patients ('small' T3 tumours, 'localized' T4 tumours) will be operable but, for most patients, their initial management will be non-surgical. The diagnosis should be confirmed by core biopsy and the receptor status measured.

Inflammatory breast cancer

This aggressive locally advanced breast cancer presents clinically as an inflamed and tender breast; the patient may well have been initially treated with antibiotics. The diagnosis should be confirmed by core biopsy, although this may need more than one attempt, since the inflammatory reaction within the breast may make it difficult to obtain definitive histology. Inflammatory breast cancer is associated with high morbidity and mortality and has a 5-year survival of 50%, with only 30% of patients disease free at 10 years.

These tumours do not respond well to endocrine management and, because of their rapid growth, should be treated with initial chemotherapy and then subsequent surgery/radiotherapy to achieve the best local control, disease-free survival and overall survival.

Overview of breast cancer management

It is well established that breast cancer is a spectrum of diseases associated with different clinical behaviours and treatment should be tailored appropriately. Therefore, management of patients with breast cancer may involve surgery, radiotherapy, systemic therapy (hormonal and/or chemotherapy) or a combination of these. Since there remain uncertainties in the management of breast cancer, consenting patients should be entered into appropriate clinical trials when possible. Many patients want to be, and benefit from being, fully informed of the possible treatment options, and involved in the decisions about their treatment. The availability of a breast care nurse and discussion within a multidisciplinary team is helpful in this respect.

Management options by stage

These are as follows:

- All patients with T1–3, N0 or 1, M0 should be considered for surgical resection of the primary tumour and evaluation/resection of the ipsilateral axillary nodes; however, surgery may follow the administration of systemic therapy (neoadjuvant therapy) in select patients.
- Some patients with larger T2 (>2 cm) or T3, N0 or 1, M0 tumours may be considered for primary systemic therapy (neoadjuvant endocrine, chemotherapy or biologically targeted therapy) prior to surgery. This neoadjuvant approach may improve rates of breast conservation and can be informative about the individual tumour response to therapy.
- Most T4 tumours are initially inoperable but may become operable after a course of primary systemic therapy or radiotherapy.
- Very small numbers of patients (usually elderly) with potentially operable breast cancer are not anaesthetically fit for a surgical procedure; if ER positive, primary endocrine therapy may be beneficial.
- Patients with established distant metastatic disease should be managed palliatively but actively. Core biopsy of metastasis, used to confirm the diagnosis, will also allow ER and HER2 assessment.

Most patients will complete their initial treatment and then enter survivorship follow-up.

Operable breast cancer

Surgery to the breast

There are three established surgical procedures for invasive breast cancers:

1. wide local excision
2. quadrantectomy or segmentectomy
3. mastectomy.

Both wide local excision and segmentectomy would normally be followed by radiotherapy. The combination of wide local excision or segmentectomy and radiotherapy is often called breast conservation. Randomized controlled clinical trials have shown that in tumours up to 4 cm in size treatment by mastectomy or breast conservation (surgery plus radiotherapy) results in no significant difference in overall survival. Local recurrence rates are similar with a non-significant relative reduction in favour of mastectomy. Patients undergoing conservation have similar psychological morbidity to those undergoing mastectomy, but have greater freedom of dress and better body image than women who have had mastectomy.

Conservation surgery

Wide local excision is excision of a tumour with a margin of clearance of both invasive and *in situ* disease. Wide local excision (sometimes referred to as lumpectomy) should be performed with due respect to the skin lines of the breast and the likely cosmetic result (Figure 17.16); for invasive disease, an axillary procedure (see below) may be conducted under the same anaesthetic. Lateral margins (histopathological) should be 1 mm or more clear of disease. There are no direct comparisons between wide local excision (1 cm macroscopic clearance) and segmental excision (1 cm macroscopic clearance but the excision incorporating tissue from the nipple right out to the periphery of the breast) or quadrantectomy (similar excision to segmental excision but with 2–3 cm macroscopic margin clearance). The most important factor related to cosmetic outcome is the volume of tissue excised relative to the volume of the breast and therefore quadrantectomies if not followed by reconstruction (e.g. a latissimus dorsi miniflap) produce less good cosmetic results than wide local excision or segmental excision.

A central tumour is not a contraindication to conservation although it may necessitate removal of the nipple and areola, which may compromise cosmesis for some patients. Similarly, oncoplastic remodelling of the breast at the time of resection can reduce the shape and volume of deformities.

Oncoplastic remodelling can range from the judicious apposition of tissues with undermining of the skin through swung flaps of breast tissue (including skin if necessary) to reduction mammoplasty approaches. The contralateral breast may require symmetrization procedures, which may be achieved under the same operation or left to a later date.

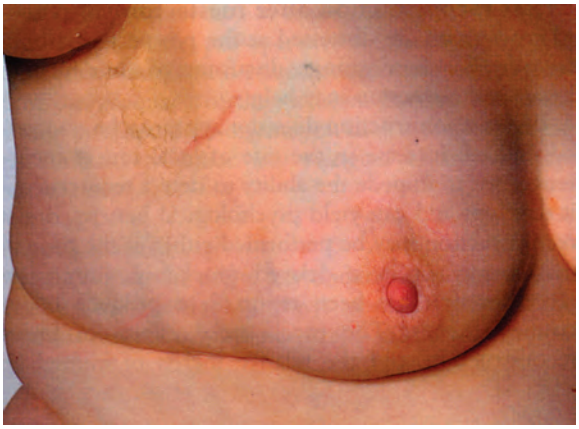


Figure 17.16 Breast conservation: surgical scars following wide local excision of a breast cancer from the 12 o'clock position and axillary node sampling.

Mastectomy

The decision whether to recommend mastectomy or breast conservation will depend on the factors listed in Box 17.6.

Mastectomy is indicated for operable breast cancer which is either large or at multiple sites, when radiotherapy is to be avoided, or by patient preference. A (now standard) modified radical mastectomy removes the skin overlying the cancer, nipple areolar complex and all breast tissue including the axillary tail, but leaves the pectoral muscle intact (removed in a radical mastectomy) and may avoid the need for radiotherapy in many cases. Patients with high-grade tumours, pathology size >5 cm, lymphatic permeation or multiple axillary node involvement have a significantly higher risk of locoregional recurrence that can be reduced by supplementing mastectomy with postoperative radiotherapy. When mastectomy is carried out, axillary clearance is usually performed.

Mastectomy may be complicated postoperatively by haematoma or seroma formation beneath the skin flaps (and hence most surgeons leave in a vacuum drain beneath the skin flaps), wound infection (which should complicate <2% of operations) or, rarely, flap necrosis.

Breast reconstruction after mastectomy

Many women are distressed at the prospect of a mastectomy. For these women, discussion of breast reconstruction may help.

Breast reconstruction does not appear to be associated with an increase in the rate of local cancer recurrence nor does it impede the ability to detect recurrence if it develops, and it can yield psychological benefit. Breast reconstruction may be performed either at the time of mastectomy or as a delayed procedure. Immediate reconstruction has been reported to

BOX 17.6 Factors influencing choice of breast surgery

- The ratio of the size of the tumour to the size of the breast:
  - smaller tumours in larger breasts are more suitable for breast conservation than larger tumours in smaller breasts: a cut-off of 4 cm tumour diameter is often quoted
- The pathological features of the tumour (if known):
  - extensive *in situ* component
  - histological grade II or III
  - lymphatic/vascular invasion
- There is an increased risk of local recurrence if disease (invasive or DCIS) is <1 mm from the margins of excision or present at multiple sites
- Age of patient:
  - Patients aged <35 years are at increased risk of local recurrence
- The patient's own preference
- Fitness for surgery and/or radiotherapy.

produce better cosmetic results. The psychosocial effects of breast reconstruction, and the relative merits of immediate and delayed surgery, have not been adequately studied.

The choice of operation for an individual patient depends on several factors including tumour site, breast size, the adequacy of skin flaps and whether radiotherapy is planned or has been previously used.

Methods comprise:

- An implant filled with silicone or saline that can be placed subcutaneously or subpectorally. Despite some adverse publicity there is no evidence that silicone prostheses are associated with significant systemic problems. Soya oil implants have now been withdrawn worldwide leaving saline-filled implants as the only alternative.
- Use of a tissue expander, placed in either the subpectoral or subcutaneous plane, which can be expanded to provide a matched size with a contralateral side and then partially deflated to allow ptosis or replaced with a silicone or saline implant with repositioning as necessary (Figure 17.17). Tissue expansion is not recommended following radiotherapy owing to the fibrotic process induced by radiation.
- Myocutaneous flaps. (1) A latissimus dorsi (LD) flap together with overlying skin based on the neurovascular supply to the LD may be used to replace a defect after mastectomy but often requires a prosthesis deep to the muscle to ensure adequate bulk and ptosis on the reconstructed side (Figure 17.18). LD miniflaps, in which skin is not taken, can be used to replace bulk following quadrantectomy and may be harvested using laparoscopic equipment and techniques. (2) Transverse rectus abdominis myocutaneous (TRAM) flap may be performed either as a pedicle flap based on the superior epigastric artery or as a free flap based on the deep inferior epigastric artery with a microvascular anastomosis between the inferior epigastric and the subscapular vessels or internal mammary artery. This provides a good bulk of tissue of similar consistency to normal breast and, together with contralateral surgery and nipple reconstruction or tattooing, can give outstanding results (Figure 17.19). The deep inferior epigastric flap (DIEP) uses perforating vessels from the epigastric artery and, although this involves microvascular anastomosis (usually to the internal mammary vessels), results in less abdominal wall defect. Other free autologous flaps include the gracilis or gluteal vessel based flaps.

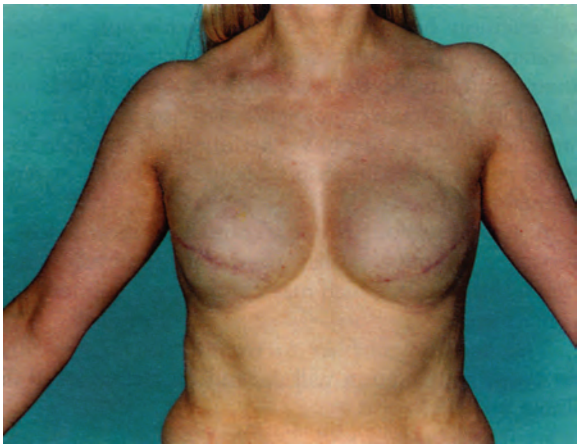


Figure 17.17 Patient who has undergone bilateral mastectomy, bilateral tissue expansion and most recently replacement of the expanders with silicone prostheses.

Surgery to the opposite breast, mastopexy or reduction mammoplasty, may be required to achieve symmetry. Techniques for reconstruction of the nipple/areola complex have been described. Alternatively, acceptable nipple prostheses may be made by taking a mould from the existing nipple or nipple tattooing.

Complications of breast reconstruction include infection of the prosthesis, necrosis of skin or flap tissue (ranging from minor fat necrosis to necrosis of the whole flap in <5% of patients)

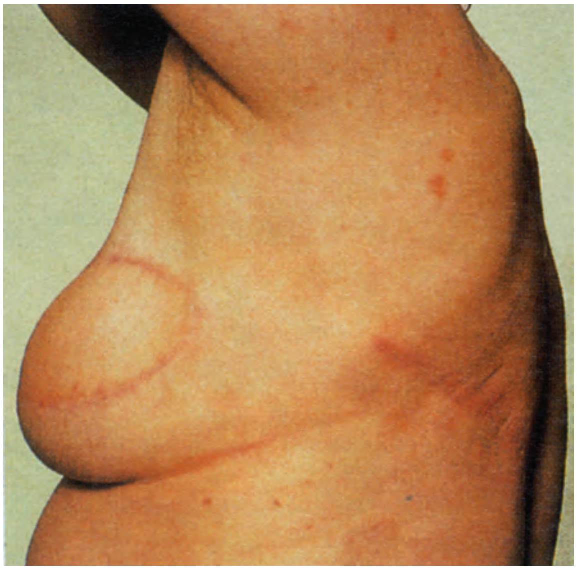


Figure 17.18 Lateral view of latissimus dorsi reconstruction showing the posterior scar and the skin/muscle reconstruction (with silicone implant) demonstrating good ptosis of the breast.

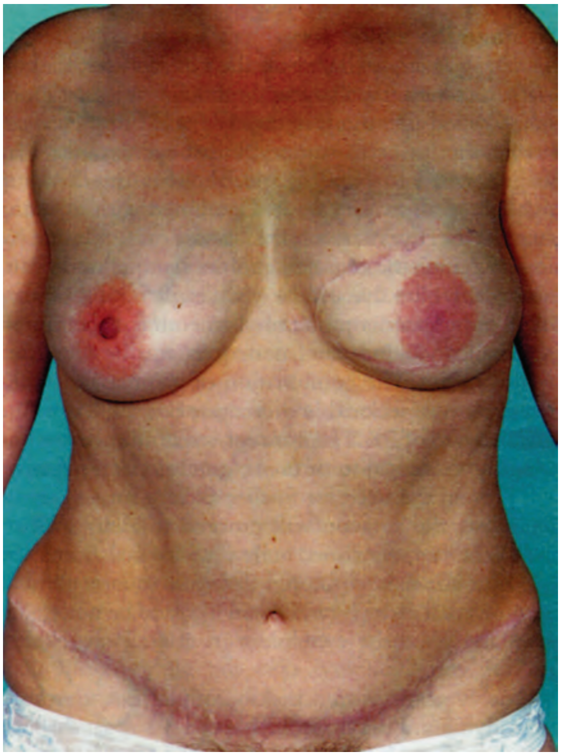


Figure 17.19 Left transverse rectus abdominis myocutaneous free flap reconstruction plus nipple tattooing. Note abdominal donorsite scar.

and puncture or leakage of implants (10% at 10 years); fibrotic capsule formation around the implant (requiring capsulotomy or capsulectomy) is now less common (<10% at 1 year) owing to the use of textured implants, but cosmetically unsatisfactory results can still occur.

Who should perform breast reconstruction? Should it be the same surgeon who performs the mastectomy? There are advantages and disadvantages to the same surgeon performing both operations. Certainly, the surgeon performing the reconstruction should be fully trained in all the appropriate techniques and, in most units, will be a plastic surgeon. Patients who are being prepared for a mastectomy should be informed of the option of reconstruction and, if appropriate, should discuss the options with a surgeon trained in reconstructive techniques, prior to their surgery.

Impalpable tumours

Mammographic screening is increasingly detecting lesions which are radiologically suspicious of malignancy but impalpable. While stereotactic core or vacuum biopsy may provide satisfactory diagnostic material, excision of an impalpable lesion may be required for diagnosis, and excision of histologically proven cancer for therapeutic breast conservation requires localization of the radiologically detected lesion. Excision of a needle-localized radiological abnormality (Figure 17.20) may thus be considered diagnostic (if lesser procedures have failed to establish the diagnosis) when the procedure will simply excise the area targeted, or a therapeutic procedure to excise the radiologically abnormal area with a 10 mm rim of normal tissue while at the same operation axillary sentinel node biopsy may be performed. The diagnostic biopsy of a benign lesion should weigh <30 g, the intention being to minimize breast disruption by the diagnostic procedure.

Localization can be performed under either ultrasound or stereotactic guidance and then guiding one or more fine wires to the localized lesion. A check film is taken to allow the surgeon to place an appropriate incision and excise the abnormal area (Figure 17.20a).

Specimen radiology of the excised lesion while the patient remains anaesthetized (local anaesthesia or general anaesthesia may be used for this procedure) ensures that the correct area of tissue has been removed by comparison with preoperative mammography (Figure 17.20b), and subsequent histological examination is required to confirm the adequacy of the excision. Excised specimens should be marked for orientation so that the pathologist can indicate whether any margins of excision are inadequate.

Paget's disease

Depending on the site of any underlying primary, either wide local excision (with/without radiotherapy) or mastectomy (if the lesion is distinct from the Paget's disease) is satisfactory.

Ductal carcinoma *in situ*

Localized areas of DCIS should be completely excised with a wide margin (minimum 2 mm) on pathology review.

In patients with extensive DCIS (>4 cm) (Figure 17.21) or disease affecting more than one quadrant, a mastectomy should usually be performed and results in 98% 5 year survival. Surgical staging of the axilla is not required for DCIS alone.

Following adequate local excision of DCIS, patients should be considered for radiotherapy to the breast. While a 1 cm margin of normal tissue around DCIS may be sufficient to gain local control, there is a reduction in recurrence or the development of invasive cancer in those given radiotherapy following wide local

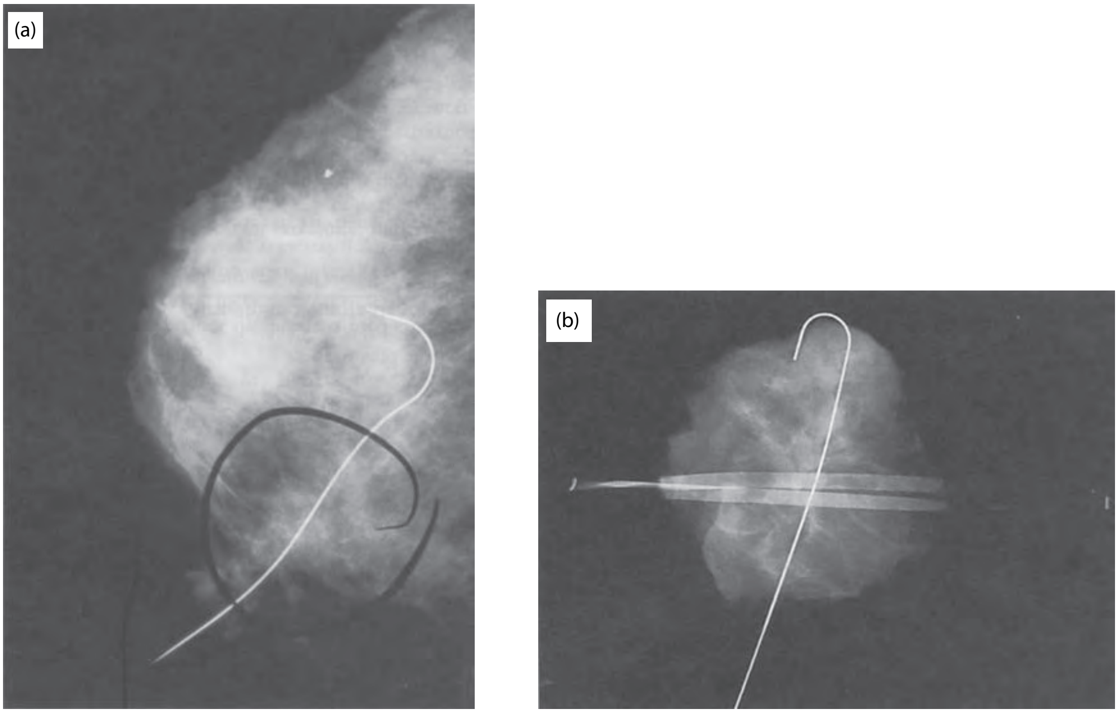


Figure 17.20 Needle localization for impalpable screen-detected cancer: (a) mammogram with localizing wire in place; (b) excision biopsy confirmatory film.

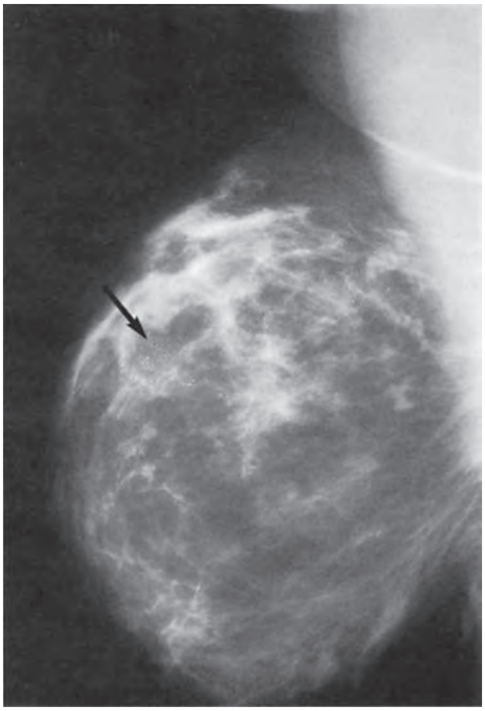


Figure 17.21 Ductal carcinoma *in situ* on a screening mammogram showing pleomorphic branching calcifications.

excision of high-grade DCIS. The role of systemic endocrine treatment may be to reduce development of contralateral DCIS.

Axillary surgery is not usually recommended in patients with DCIS alone; however, patients with large areas of high-grade DCIS have a 50% chance of occult invasion on excision. In these circumstances, sentinel node biopsy may be appropriate.

Lobular carcinoma *in situ*

Women with LCIS diagnosed on core biopsy require surgical diagnostic biopsy to exclude the presence of DCIS or invasive cancer. Women with LCIS diagnosed or confirmed at surgery require mammographic screening because of their increased risk of future breast cancer.

Surgery to the axilla

Lymphatic drainage from the breast is predominantly to the axilla, although metastatic breast cancer may spread via internal mammary, intercostal or interpectoral nodes.

One of the most important prognostic indicators is whether there has been spread of cancer to the axillary lymph nodes (Figure 17.15), and hence it is used as one of the major determinants of appropriate systemic adjuvant therapy. Axillary surgery should be performed in all patients with invasive operable breast cancer. Axillary nodes lie inferior to the axillary vein and are classified in relation to the pectoralis minor muscle, with level I lateral to, level II posterior to and level III medial to pectoralis minor.

Although present in only 23% of screen-detected cancers, some 40% of symptomatic cancers have axillary metastases.

For larger tumours (T3 and above) a complete level I, II and III axillary clearance is the treatment of choice.

Table 17.4 Recommended management of axillary nodes for operable breast cancer

Impalpable cancer or <4 cm, clinically node negative	Sentinel node biopsy
Proven axillary metastases	Level I, II and III clearance

Since only 5% of node metastases are ‘skip metastases’, missing the lower level I nodes but involving level II or III, for small cancers there is emerging consensus as to the best way to manage the axilla (Table 17.4).

The following procedures are practised.

To stage the axilla

Sentinel node biopsy

Based on the concept that there is a first or sentinel axillary node draining breast cancer, radiolabelled albumin is injected adjacent to the cancer up to 24 hours prior to surgery; breast scintigraphy may be performed to confirm that the label has travelled and highlighted an axillary node. Combined with injecting patent blue dye around the cancer at the time of surgery, lymphatic channels and the sentinel node containing the blue dye can be visually identified (Figure 17.22) and a gamma probe used to confirm the hottest (sentinel) node. A second node may commonly be identified by this means. The combination of gamma and visual detection is superior to either alone but still has a false-negative rate of 5%. Immediate histological examination of the sentinel node, if negative, will prevent the need for further axillary surgery. If positive, by cytology, frozen section or polymerase chain reaction at the time of surgery or on subsequent immunohistochemistry, the surgeon will proceed to immediate or subsequent, axillary clearance.

To treat the axilla

Axillary node clearance – a block dissection of the axillary contents – comprises the tissue bounded by the axillary vein (superiorly), LD, serratus anterior and pectoralis major, and the apex is situated at the point where the axillary vein passes over the first rib.

Level I is up to the lateral border of pectoralis minor; level II is up to the medial border of pectoralis minor; and level III is up to

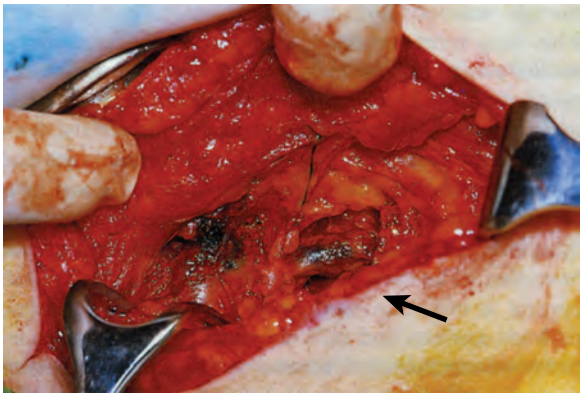


Figure 17.22 Sentinel node biopsy: blue dye staining lymphatics and two sentinel nodes. The lower node contained a high radiolabel signal and had a small focus of metastatic breast cancer.

the apex of the axilla. Care is taken to preserve the neurovascular supply to serratus, latissimus and pectoralis major and cutaneous nerves and not to dissect superior to the axillary vein.

Complications of axillary clearance include seroma formation (50%), infection (5%), cutaneous nerve damage, shoulder stiffness (5%) and pain. Long-term complications include upper limb lymphoedema (up to 15%), neural damage or restricted shoulder movement (10%).

There is no advantage of axillary clearance over axillary radiation in terms of survival at 10 years, nor does either have an advantage in terms of regional control of disease.

Only level III dissection fully stages the axilla and treats nodal disease. Further treatment of the axilla (radiotherapy or surgical re-excision) may be required for patients who have had a level I or II dissection to achieve local disease control if involved nodes are identified histologically, but risks lymphoedema. Further treatment after level III dissection is only considered if the nodes cannot be adequately cleared or if there is extranodal spread of the tumour. Clearance and radiotherapy increases the risk of lymphoedema of the upper limb to some 40%.

Only 5% of women with breast cancer have internal mammary node metastases, usually from medially placed cancers, but 90% of these women also have axillary metastases, so most surgeons do not perform internal mammary node biopsy.

Rarely, breast cancer may present with axillary nodes but no clinically evident breast primary. While mammography and MRI may identify an occult lesion in two-thirds of these patients, in the remainder level III axillary clearance with observation (clinical and radiological) of the breast may be the treatment of choice.

Adjuvant radiotherapy

Radiotherapy has been used after surgery for breast cancer for decades. Indeed the move from radical mastectomy was facilitated by trials showing that simple mastectomy and radiotherapy produced results equivalent to radical mastectomy, with less loss of function. A generation later, trials showed that, for patients with small operable tumours, wide local excision and postoperative radiotherapy were equivalent in terms of survival and local control to conventional mastectomy. Postoperative radiotherapy is now used following mastectomy (selectively) or breast conservation for operable breast cancer.

Postmastectomy radiotherapy

Radiotherapy and local control

Radiotherapy given after mastectomy will reduce the risk of isolated local recurrence by approximately two-thirds. The absolute benefit will depend on the risk of recurrence: thus, if the risk of recurrence without radiotherapy is 39%, this will be reduced to 13% with radiotherapy, an absolute benefit of 39% − 13% = 26%. Conversely, a risk of recurrence of 3% would be reduced to 1%, an absolute benefit of 2%. The former would certainly be clinically useful, whereas, for the latter, the toxicities of radiotherapy would probably outweigh the (modest) benefit.

Radiotherapy after mastectomy is thus generally recommended only for those patients judged to be at a particular risk of recurrence, i.e. large primary tumour size (>5 cm), more than three involved nodes, high histological grade, lymphatic and/or vascular invasion and involved (usually deep) margins. The risk of local recurrence is a summation of these factors, although there is no convenient formula to quantify them.

Systemic adjuvant therapy (see page 398) does reduce the risk of local recurrence, but radiotherapy confers additional benefit and the relative reduction in local recurrence is independent of any systemic therapy. Thus, the decision as to whether to recommend postmastectomy radiotherapy depends on the perceived risk of recurrence, bearing in mind the pathological factors considered above and the systemic therapy to be offered.

Extent of radiotherapy

For the reasons mentioned above, radiotherapy will normally certainly include the chest wall. Whether the axilla is included depends on the type and results of axillary surgery. If the axilla has been formally assessed as negative by sentinel node biopsy then axillary radiotherapy will not improve local control within the axilla. Conversely, a positive sample should be followed by either level III axillary clearance or axillary radiotherapy. A surgical clearance of the axilla should both stage and treat the axilla. Radiotherapy following axillary clearance will substantially increase the risk of morbidity without any increase in local control. The only exceptions are when a clearance is known to have left residual disease and/or there is extensive extracapsular spread – these patients are at high risk of local recurrence with its associated morbidity, which must be balanced against the morbidity of treatment.

The supraclavicular fossa and internal mammary chain of nodes have been irradiated as part of standard therapy in some centres, but without clear evidence of benefit. Isolated internal mammary chain recurrence is rare and supraclavicular recurrence is a manifestation of metastatic relapse.

Radiotherapy after breast conservation

All studies have shown that radiotherapy to the breast following limited breast surgery reduces the risk of local recurrence. Table 17.5 shows the recurrence rate for three trials in which different surgical procedures were used. Although these are not direct comparisons, the data suggest that, when more extensive surgery (quandrantectomy or sector resection) is used, the local recurrence

Table 17.5 Comparison of effect of radiotherapy and the extent of breast surgery on recurrence rate

Type of surgery	**Wide local excision**	**Sector resection**	**Quadrantectomy**
Name of trial	NSABP	Uppsala–Orebro	Milan
No. of patients	1450	381	579
No radiotherapy	39%	7.6%	10%
Radiotherapy	10%	2.9%	0%
Duration	8 years	3 years	3 years

NSABP, National Surgical Adjuvant Breast and Bowel Project.

rate without radiotherapy is less (albeit at the cost of an impaired cosmetic result). Nevertheless, even with more extensive surgery, radiotherapy significantly reduces the local recurrence rate.

If the patient has systemic adjuvant therapy as well as surgery, radiotherapy further reduces the risk of local recurrence; the magnitude of the effect is very similar to that seen after mastectomy. Trials have not identified a group of patients in whom radiotherapy does not reduce the risk of recurrence and hence in whom it could be safely omitted. Nevertheless, the absolute benefit of radiotherapy for (especially older) patients with small, grade 1, node-negative breast cancers is small and there are ongoing trials quantifying this risk. Thus, radiotherapy is of potential benefit to all patients undergoing breast conservation and should only be omitted if the clinician feels, after discussion with the patient, that the morbidity of radiotherapy does not justify the excess recurrence risk for that patient.

Most recurrences after breast-conserving surgery occur in the tumour bed and hence some studies have looked at whether radiotherapy to the tumour bed alone may be adequate in terms of local control. This has consisted of a single dose of intraoperative radiotherapy given at the time of surgery (the TARGIT trial) or fractionated brachytherapy given later (ELIOT trial).

Effect on survival

Although the importance of local treatment (surgery, radiotherapy) on local control has long been recognized, the concept of breast cancer as a systemic disease at presentation used to discount the importance of local therapy in terms of survival. An overview of trials of radiotherapy (after both breast-conserving surgery and mastectomy) has shown that an improvement in local control at 5 years translates into an improvement in 15 year mortality – approximately for every four local recurrences avoided, there is an avoidance of one breast cancer death at 15 years. There is a similar but smaller effect on 15 year overall mortality. Longer follow-up (>15 years) has also shown a modest but real excess in overall mortality with an excess of deaths from heart disease and lung cancer in irradiated patients. These trials were inevitably started decades ago (1960s and 1970s), when radiotherapy techniques were very different, so one would expect (and hope) that modern techniques that limit the amount of heart and lung irradiated would be safer. There is thus good evidence that improving local control can affect survival, presumably by the recurrence acting as the focus for metastatic reseeding. The practical implications are that local control is important and needs meticulous technique by both surgeon and radiotherapist to minimize morbidity, maximize local control and hence improve survival.

Dosage and complications

Radiotherapy was traditionally given to the breast/chest wall and lymphatics as a dose of 50 Gy in 25 daily fractions over 5 weeks, but recent trials have shown that equivalent rates of local control and cosmesis can be obtained by giving 40–42 Gy in 15 or 16 fractions over 3 weeks, as exemplified by the START trial. Further reductions in the number of fractions ('hypofractionation') are the subject of ongoing studies.

Since local recurrences occur in the tumour bed, it has been common practice to give a higher dose (a 'boost') to the tumour bed. A large EORTC trial showed that this improved local control but that this was at the expense of some impairment in cosmesis. The benefit of the boost was mainly in patients at higher risk of recurrence (mainly younger patients, i.e. under 50), and hence a policy of tailoring the boost to the risk of recurrence is widely practised.

Patients can expect a degree of skin erythema and fatigue during and shortly after radiotherapy to the breast. Supraclavicular radiotherapy may cause transient oesophagitis. In the long term, a small number of patients develop breast oedema and/or fibrosis. As mentioned earlier, radiotherapy to the axilla can cause lymphoedema, especially if given after axillary clearance.

Systemic therapy for breast cancer

Breast cancer was the first cancer to be treated systemically when Beatson reported the benefits of oophorectomy for young patients with advanced breast cancer in 1896. Most patients will receive some form of systemic therapy, whether this be adjuvant or palliative in intent.

Adjuvant systemic therapy

For many patients, invasive breast cancer is a systemic disease at presentation, albeit the metastases are not detectable and only become apparent during the ensuing years. The rationale of systemic adjuvant therapy given after surgery is to treat and hopefully eradicate the occult distant metastases and hence improve survival. There have been numerous trials of adjuvant therapy and the results of these have contributed to the regular overviews of the Early Breast Cancer Trialists' Collaborative Group (EBCTCG). These confirm that adjuvant endocrine, cytotoxic and biological therapy can improve both recurrence-free and overall survival in a subset of patients with clinically localized disease. The principles and magnitude of the benefits of adjuvant therapy can be illustrated by considering the results of trials of ovarian ablation. For premenopausal women not receiving chemotherapy, ovarian ablation reduces the odds of death by 24%, but Figure 17.23 shows, first, that there is a slightly greater effect on recurrence-free survival than on overall survival, and, second, that the absolute reduction varies according to the risk of relapse. In other words, node-positive women have a higher absolute benefit because they are at a higher risk of relapse.

In effect there are three groups of women, as summarized in Table 17.6 (percentage defined at 10 years).

Women in group A have significant occult metastatic disease that is either resistant to or not eradicated by the systemic therapy. They may gain some prolongation of disease-free survival but they will not gain any survival advantage.

Those in group B have occult distant metastases that are eradicated by systemic therapy and hence achieve a personal cure; they are thus the true beneficiaries of systemic adjuvant therapy (and represent the difference between the curves in Figure 17.23).

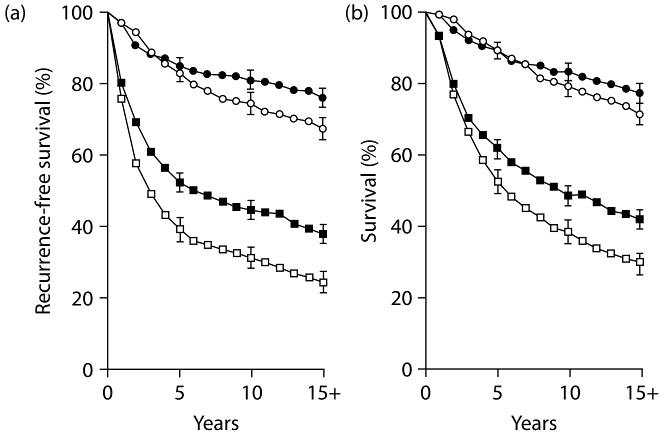


Figure 17.23 Effects of ovarian ablation alone in breast cancer.
(a) Recurrence-free survival for node-negative women <50 years (upper pair of curves), demonstrating improved survival for women who had ovarian ablation (top line) versus controls (second line); for node-positive women <50 years (lower pair of curves), better survival for patients who received ovarian ablation (third line down) versus control (bottom line). (b) Overall survival for node-positive women <50 years (upper pair of curves), demonstrating improved survival for women who had ovarian ablation (top line) versus controls (second line); for node-positive women <50 years (lower pair of curves), better survival for patients who received ovarian ablation (third line down) versus control (bottom line).

Table 17.6 Relative benefit of ovarian ablation on survival (see text)

Group	A	B	C	Total
Node negative	24%	6%	70%	100%
Node positive	58%	13%	29%	100%
Occult metastases	Resistant	Sensitive	None	

Those in group C had no metastases at presentation and do not benefit from systemic therapy since they would have survived anyway.

Table 17.6 also summarizes the relative size of the groups for both node-positive and -negative patients. Adjuvant systemic therapy can significantly improve the outlook for patients with breast cancer and has contributed to the recent improvement in survival. Nevertheless, most patients will not derive personal benefit from it and it is important to balance the morbidity of treatment (for all) against the benefit (for a minority).

Although a survival benefit has rarely been demonstrated with the administration of therapy prior to surgical resection, neoadjuvant therapy should be strongly considered in a subset of patients with localized breast cancer. In these cases, the diagnosis should be confirmed by a core biopsy, since a positive fine needle aspirate does not discriminate between invasive and *in situ* disease.

Specifically, candidates for neoadjuvant therapy include patients with locally advanced breast cancer, patients with triple receptor-negative breast cancer >1–2 cm in size or with positive lymph nodes on initial evaluation, young patients with tumours >2 cm in size regardless of receptor or nodal status, or patients with large tumours who are interested in breast-conserving surgery. The administration of neoadjuvant therapy has no worse outcome than the administration of adjuvant therapy; however, several benefits can be derived from preoperative systemic therapy. First, by decreasing the size of the primary tumour, tumours may be 'downstaged' from inoperable to operable in select patients. Second, although most patients (~80%) will have either a complete (CR) or partial response in the primary tumour, because the effects of treatment are directly observed, therapy can be altered in cases of disease progression, which results in less patient exposure to the toxicity of ineffective drugs. Additionally, a subset of patients will develop a complete pathological response (pCR) to neoadjuvant therapy, which has been associated with an extremely good prognosis. Finally, the use of neoadjuvant therapy has become a powerful tool in drug development as it offers the opportunity to obtain tissue before and after therapy to correlate tumour response with molecular parameters of drug response.

Choice of systemic therapy

The choice of appropriate adjuvant (and neoadjuvant) therapy will be determined by the:

- receptor status of the tumour
- patient's menopausal status
- patient's risk of recurrence (based mainly on age, nodal status, tumour size and tumour grade).

The selection of systemic therapy requires individual discussion of clinical and pathology findings, which may be helped by using the NPI, Adjuvant! Online or similar programs, and molecular testing as an aid to decision-making (see page 385).

Endocrine therapy

In most women, defining menopausal status is usually straightforward. Particular caution should, however, be exercised in premenopausal women whose periods stop after adjuvant chemotherapy. Their periods can return some considerable time after chemotherapy and hence such women should not be treated with aromatase inhibitors.

Ovarian ablation

Ovarian ablation reduces the odds of recurrence by 25% and of death by 24% (a similar order of improvement to that achieved by polychemotherapy), but is only of benefit to premenopausal women. Many of the trials of ovarian ablation were carried out before hormone receptor status was measured, but more recent studies suggest that the benefit is clearly confined to women with hormone receptor-positive tumours. Ovarian ablation can be performed either laparoscopically, by a short course (usually four fractions) of low-dose radiotherapy to the pelvis ('radiation menopause'), or medically using a gonadotrophin analogue (usually given as a monthly subcutaneous injection into the abdominal wall). The optimal duration of 'medical oophorectomy' has not been established, but has generally been 2–5 years and has become the favoured method because of its reversibility. All methods seem to work, although there are few comparative data. The main side effects of all approaches are those of a menopause – flushes, sweats, mood changes – but are often worse than a physiological menopause, because of the precipitative nature of the menopause.

Selective oestrogen response modulators

Selective oestrogen response modulators (SERMs) structurally resemble oestrogens and bind to ERs, which decreases proliferation and induces tumour regression in ER-positive breast cancer cells. The most commonly prescribed SERM is tamoxifen, a drug which has demonstrated clinical benefit for the treatment of hormone receptor-positive breast cancer in both the adjuvant and metastatic settings. Although tamoxifen acts as an 'antioestrogen' in cancer cells, it is an oestrogen agonist in normal tissue. For example, in postmenopausal women, tamoxifen gives some protection against osteoporosis and reduces blood cholesterol. Additionally, tamoxifen is associated with a small, but significant, increased risk of venous thrombotic disease and endometrial hyperplasia, which can progress to endometrial cancer.

In the adjuvant setting, 5 years of tamoxifen was associated with a 41% reduction in the risk of recurrence and 34% reduction in the risk of death from breast cancer in women with ER-positive breast cancer. This effect was greater than that seen with 2 years of adjuvant tamoxifen. Ten years of adjuvant tamoxifen is better than 5 years at reducing the risk of breast cancer recurrence, although the effect is seen only beyond 10 years, thus establishing 10 years of tamoxifen therapy as the standard of care for the adjuvant treatment of hormone receptor-positive breast cancer in premenopausal women.

Tamoxifen is reasonably well tolerated, but up to 40% of women can experience side effects such as hot flushes, weight gain, gastrointestinal upset, loss of libido and vaginal dryness or discharge. For some women, the severity of these side effects may be enough for them to cease taking tamoxifen.

Aromatase inhibitors

For postmenopausal women, oestrogen is produced by aromatization of androgenic precursors. Inhibition of aromatase, the enzyme that converts androgenic substrates into oestrogens, has been a target for drug development for over two decades; however, early inhibitors were associated with significant toxicity and incomplete suppression of oestrogen synthesis. Third-generation aromatase inhibitors, such as anastrozole, letrozole and exemestane, were noted to have improved efficacy with acceptable toxicity and are commonly used to treat postmenopausal women with either localized or advanced hormone receptor-positive breast cancer.

In the adjuvant setting, aromatase inhibitors have also demonstrated improved efficacy compared with tamoxifen in risk reduction for breast cancer recurrence. The Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial was a randomized double-blind adjuvant therapy trial conducted in postmenopausal women with surgically treated, hormone receptor-positive (or unknown) breast cancer. Patients were randomized to receive either oral anastrozole 1 mg daily or oral tamoxifen 20 mg daily, or the combination. Interim analysis of the combination arm demonstrated no improvement compared with tamoxifen alone and accrual to this arm was halted. Single agent anastrozole significantly improved disease-free survival and time to recurrence compared with single agent tamoxifen. Patients receiving anastrozole also developed fewer contralateral breast primary tumours than those taking tamoxifen.

These results have been validated by additional studies with letrozole or exemestane as primary adjuvant therapy or following tamoxifen in a cross-over design. To date, it is not clearly established whether initial therapy with an aromatase inhibitor is preferred to therapy with tamoxifen followed by an aromatase inhibitor. Currently, the use of adjuvant therapy with aromatase inhibitors is limited to postmenopausal patients; however, data from ongoing clinical trials such as the SOFT or TEXT trials are maturing and will determine the benefit of ovarian ablation and aromatase inhibition as adjuvant therapy in premenopausal women with hormone receptor-positive breast cancer.

Neoadjuvant endocrine therapy has also been used for the treatment of postmenopausal women with operable hormone receptor-positive breast cancer and, in at least one small randomized phase II study in postmenopausal hormone receptor-positive breast cancer, treatment with the aromatase inhibitor exemestane demonstrated similar response rates to doxorubicin and paclitaxel. Several randomized trials have demonstrated improved response rates for neoadjuvant therapy with aromatase inhibitors compared with tamoxifen. There appears to be similar efficacy among the aromatase inhibitors used for neoadjuvant therapy.

Cytotoxic therapy

Because of toxicity and the risk of long-term morbidity, the use of chemotherapy is a complex decision in the treatment of breast cancer. In the adjuvant and neoadjuvant settings, the use of chemotherapy depends upon the risk of recurrence and the receptor status of the tumour.

Many chemotherapy drugs have activity against breast cancer. Anthracyclines exert their effects by inhibiting topoisomerase II and increasing DNA fragmentation. Taxanes, epothilones, vinca alkaloids and analogues of halichondrin B exert their cytotoxicity by affecting microtubular dynamics and blocking the progression of mitosis. Nucleoside analogues, such as 5-fluorouracil, capecitabine and gemcitabine, induce apoptosis by affecting DNA replication. Alkylating agents (e.g. cyclophosphamide) and platinum salts cross-link DNA strands and induce apoptosis. Common toxicities associated with chemotherapy drugs used to treat breast cancer include myelosuppression, alopecia, nausea/vomiting, neuropathy and, in the case of anthracyclines, cardiomyopathy, which can cause congestive heart failure. Neutropenic sepsis is an uncommon, but serious (and occasionaly fatal), complication of chemotherapy administration and warrants immediate intervention with broad-spectrum antibiotics, including coverage for Gram-negative pathogens. Chemotherapy can also induce cessation of menses, which may be permanent; thus, premenopausal patients should always be warned of the risk of an early menopause and loss of fertility. Not all premenopausal patients experience menopause, however, so these patients should be counselled to avoid conception, since all chemotherapy drugs can be damaging to the fetus if administered during the first trimester of pregnancy.

The EBCTCG (2012) overview confirms that adjuvant polychemotherapy using an anthracycline-containing combination, such as FEC (5-fluorouracil, epirubicin, cyclophosphamide) or an anthracycline/taxane combination, produces an approximately one-third reduction in breast cancer mortality. This proportional risk reduction is little affected by nodal status, hormone receptor status, tumour grade or size, age or the use of adjuvant tamoxifen. The absolute benefit however depends on the risk of relapse without chemotherapy. The absolute benefit of adjuvant chemotherapy is greater for node-positive women than those who are node negative because their risk of relapse is higher. Chemotherapy seems to be of benefit to patients with both ER-positive and -negative tumours.

In the adjuvant and neoadjuvant settings, a combination of drugs is more effective than one alone, and a prolonged course of multiple doses more effective than a single dose for increasing disease-free recurrence and overall survival (Table 17.7). Early adjuvant chemotherapy trials in patients with lymph node-positive disease established a regimen of cyclophosphamide, methotrexate and 5-fluorouracil (CMF) as the standard of care. Subsequent studies have incorporated an anthracycline (usually doxorubicin or epirubicin), and the overview shows that these drugs can produce a further reduction in the odds of recurrence or death (compared with CMF) of about 11–12%, which translates to an absolute survival gain of about 2–3%. Some of these studies also demonstrated similar relative risk reductions in patients with lymph node-negative breast cancer. The addition of taxanes to adjuvant anthracycline-based regimens resulted in an additional improvement in disease-free survival of ~14% and overall survival of ~13%. To date, no adjuvant trials have demonstrated improvement in survival with chemotherapy agents added to anthracycline/taxane-based regimens or the use of high-dose chemotherapy with stem cell transplantation.

Randomized controlled trials have also been conducted to evaluate the benefit of neoadjuvant compared with adjuvant chemotherapy in patients with operable breast cancer. These studies demonstrated improvements in the numbers of patients able to undergo breast-conserving surgery; however, no difference was seen in disease-free or overall survival. pCR rates can be achieved in 12–80% of women treated with neoadjuvant chemotherapy (depending on tumour subtype and drugs used) and these patients have a better prognosis with improved overall survival than those whose tumours do not respond completely to therapy.

Biological therapies

HER2 overexpression occurs in approximately 13% of primary invasive breast cancers (and some 25% of metastatic breast cancers) and has been correlated with poor prognosis. Trastuzumab, a recombinant humanized monoclonal antibody against the extracellular domain of HER2, was the first biological therapy effective in breast cancer.

Trastuzumab is considered to be standard of care as adjuvant therapy for patients with HER2-positive localized breast cancer. The National Surgical Adjuvant Breast and Bowel Project (NSABP) B-31 trial and the North Central Cancer Treatment Group N9831 intergroup trial demonstrated that the addition of trastuzumab to standard adjuvant chemotherapy (doxorubicin and cyclophosphamide every 3 weeks for four cycles followed by paclitaxel with/without trastuzumab) improved disease-free survival and overall survival in patients with localized breast cancer. Comparison of the concurrent with the sequential arms in N9831 revealed an improvement in disease-free survival favouring the concurrent arm (hazard ratio 0.75, 95% confidence interval 0.60–0.94, $p = 0.019$); however, no difference in overall survival has been reported to date. Similar results were seen in the Herceptin Adjuvant Breast International Group 01-01 trial. In this study, adjuvant trastuzumab was administered after completion of adjuvant systemic chemotherapy using a 3-weekly administration schedule of trastuzumab for 1 or 2 years. The optimal duration of trastuzumab administration (1 versus 2 years) has yet to be defined.

The Breast Cancer International Research Group 006 compared anthracycline- and non-anthracycline-based chemotherapy regimens in combination with trastuzumab. This study demonstrated improvement in disease-free survival with the addition of trastuzumab to adjuvant chemotherapy and to date there has been no significant difference in disease-

Table 17.7 Pivotal adjuvant chemotherapy trials

Author/trial	Treatment regimens	*N*	Disease-free survival	*P*-value	Overall survival	*P*-value
Bonadonna	CMF × 6 cycles vs no chemotherapy	207 179	29% 22%	NS	25% 16%	NS
NSABP B-15	AC × 4 cycles vs CMF × 6 cycles vs AC × 4 cycles→CMF × 3 cycles	734 732 728	62% 63% 68%	NS NS NS	83% 82%	NS
CALGB 9344	AC × 4 cycles vs AC × 4 cycles→paclitaxel × 4 cycles	1551 1570	65% 70%	0.0023	77% 80%	0.0064
NSABP B-28	AC × 4 cycles vs AC × 4 cycles→paclitaxel × 4 cycles	1529 1531	81% 81%	NS	92% 90%	NS

AC, doxorubicin, cyclophosphamide; CMF, cyclophosphamide, methotrexate, 5-fluorouracil; NS, not significant NSABP, National Surgical Adjuvant Breast and Bowel Project; CALGB, Cancer and Leukemia Group.

free survival in the anthracycline-containing arm compared with the non-anthracycline arm; however, it is important to note that these two arms were not statistically powered to assess equivalence. A Finnish trial that was designed to evaluate the effects of adjuvant docetaxel versus vinorelbine followed by 5-fluorouracil, epirubicin and cyclophosphamide for three cycles included a small subset of 232 patients with HER2-positive disease who also underwent randomization to either chemotherapy alone or in combination with nine treatments of concurrent, weekly trastuzumab. The magnitude of improvement in disease-free survival in this small subset was similar to that seen in trials investigating a more prolonged administration of trastuzumab, bringing into question the optimal duration of trastuzumab therapy. Trials of different durations of trastuzumab are ongoing. Trastuzumab has also demonstrated pCR rates from 12% to 65% when given in combination with neoadjuvant chemotherapy.

In general, single agent trastuzumab is well tolerated and trastuzumab given in combination with chemotherapy does not significantly increase chemotherapy toxicity, with the exception of cardiac toxicity in combination with anthracyclines. In the adjuvant trials, predetermined criteria for holding or stopping trastuzumab were implemented and the rate of congestive heart failure was approximately 4% across the adjuvant trials.

Adjuvant therapy for ductal carcinoma *in situ* and chemoprevention

Tamoxifen has been used as adjuvant therapy for the treatment of DCIS or for prevention of cancer in women at high risk for developing breast cancer. The NSABP B-24 trial randomized 1804 women who had undergone surgical excision of DCIS to receive either tamoxifen or placebo and demonstrated that 5 years of tamoxifen following breast-conserving surgery and radiation therapy for DCIS resulted in an improvement in ipsilateral breast cancer events from 10% to 8.5% (risk reduction 32%) at 15 year follow-up. This benefit was not confirmed in a second European study, which randomized 1701 patients to tamoxifen or placebo and found no difference in breast cancer events between the two arms. The NSABP P-1 study investigated the role of tamoxifen in preventing invasive breast cancer for high-risk women (*n*=13 388). This study was stopped before reaching full accrual as an interim efficacy analysis found that tamoxifen reduced the relative risk of invasive breast cancer events by 43% compared with placebo (absolute risk reduction 1.8%). A second prevention study, IBIS-I, enrolled 7152 women who were randomly assigned to receive tamoxifen or placebo. At a median follow-up of 96 months, the estimated absolute reduction in cumulative incidence of breast cancer was similar to that seen in the P-1 study. Two smaller European studies failed to show statistical benefit; however, a meta-analysis of all four trials demonstrated that tamoxifen significantly reduced the incidence of invasive breast cancer by 38% (1.7% versus 2.6%), but the incidence of serious adverse events such as endometrial cancer or thromboembolic events were also increased by tamoxifen (1.2% versus 0.6%). The STAR trial compared the use of tamoxifen with raloxifen, a SERM that lacks the oestrogen agonist effects of tamoxifen, for the prevention of breast cancer in high-risk women. This study demonstrated a similar risk reduction between the two drugs with fewer thromboembolic and endometrial proliferative/cancer events in women treated with raloxifen. To date, the use of SERMs to prevent breast cancer in high-risk patients and patients with DCIS remains controversial and it is important to note that none of the studies for prevention or treatment of non-invasive breast cancer have demonstrated an improvement in overall survival with the use of SERMs.

Primary medical therapy

Elderly fit patients (>70 years), who constitute 40% of women with potentially operable tumours, have in the past been managed with tamoxifen alone. Although there is no adverse effect on survival, two randomized controlled trials have shown that tamoxifen alone is associated with a higher rate of local recurrence than surgery and adjuvant tamoxifen. Thus, elderly patients with potentially operable tumours able to undergo surgery should be managed in the same way as younger women, rather than by using endocrine therapy as sole therapy. Some elderly women with larger cancers, inoperable disease or who are unfit for anaesthetic with a tumour that expresses ER may respond to neoadjuvant therapy with aromatase inhibitors or tamoxifen to render the tumour operable or gain control of disease for a sustained time without the option of resection. In consideration of which therapy to use, it should be noted that letrozole given for 4–12 months downstages breast cancer more significantly than tamoxifen.

Breast cancer in pregnancy

About 1% of breast cancers occur during pregnancy or lactation, and are thus diagnosed in 1:10 000 pregnancies. The difficulties of detecting a lump in the enlarging breast result in later detection (two-thirds have nodal metastasis). In the first 6 months of pregnancy, mastectomy and clearance form the mainstay of treatment (termination of pregnancy may be considered). In the third trimester, early delivery then conventional surgical or neoadjuvant anthracycline-based therapy may be used. Taxanes should be avoided at all times during pregnancy. Pregnancy after diagnosis/treatment for breast cancer does not appear to reduce survival.

Patients who are unfit for surgery

These will generally be elderly patients and, because of their comorbidity, the principles of their management will be closer to those applicable to metastatic disease than to those applicable to operable breast cancer. In general, the aim of treatment is to control the primary tumour with the least morbidity, maintaining quality of life, so letrozole is usually the most appropriate therapy when the tumour is ER positive on core biopsy.

Radiotherapy alone is not recommended for local control of advanced tumours, although local control rates of 20–40% at 5 years may be obtained in selected cases after using at least 60 Gy conventionally fractionated.

Local and regional recurrence

Local and regional recurrence is defined as recurrence in the treated breast or chest wall, axilla and previously the ipsilateral supraclavicular fossa, although the last is now regarded as a distant relapse.

Relapse in the chest wall, axilla or supraclavicular fossa (Figure 17.24) is usually found clinically – 80% occur within 2 years and most occur within 5 years of initial surgery – and is frequently a harbinger of distant spread. Distant metastases will be found at the time of local relapse in about 20% of patients and about a further 70% will develop distant metastases within 5–10 years following locoregional relapse. Local relapse should be confirmed histologically, by biopsy, since 17% of patients lose (or more rarely gain) ER and PgR positivity or gain (usually) HER2 positivity. Patients with local relapse should be restaged (full blood count, blood biochemistry, CT of chest, abdomen and pelvis), and those with distant metastases managed appropriately.

For patients with apparently isolated local recurrence, the best predictor of prognosis is the disease-free interval (time from initial treatment to local relapse) – the longer the interval the better the prognosis.

Management of local recurrence

Local recurrence may be classified as single spot relapse, multiple spot relapse or a field change.

The aim of treatment is to regain local control – uncontrolled local disease is miserable for the patient and difficult to manage. The management strategy will be determined by the previous treatment, the site of recurrence, its operability and hormone receptor status. The patient should be assessed jointly by the multidisciplinary team and treatment individualized. In general, recurrence in the chest wall or axilla should be removed surgically if feasible, as it usually is for single site or sometimes multiple site recurrence. This may be easier if it is first reduced in extent by systemic therapy (hormone therapy or chemotherapy) and, if not previously used, radiotherapy, which should be employed for field change recurrence. Systemic therapy should be reviewed and may be changed depending on the patient's condition, previous systemic adjuvant therapy and the hormone receptor status of the tumour. Intra-arterial chemotherapy, infusional 5-fluorouracil or photodynamic therapy have all been used to try to achieve local control with limited success. Tissue necrosis with infection may become a significant problem requiring topical metronidazole gel, charcoal dressings or systemic antibiotics. Progression of local disease recurrence often causes troublesome bleeding and may progress to carcinoma *en cuirasse* encircling and constricting the chest wall.

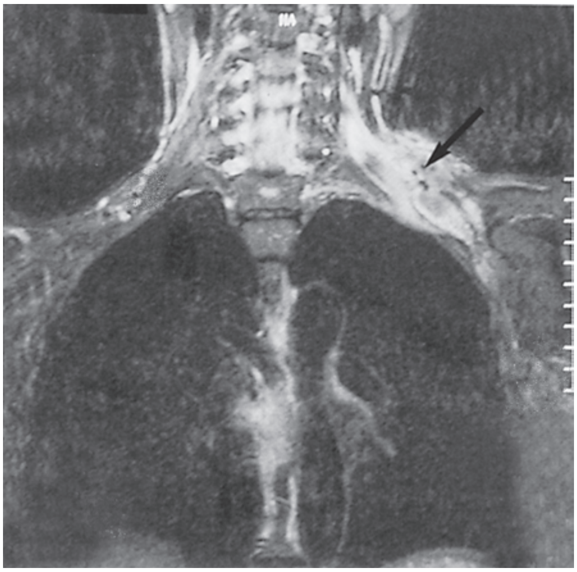


Figure 17.24 MRI showing tumour surrounding and infiltrating the left brachial plexus and right-sided pleural metastases.

Local recurrence in the treated breast

Unlike recurrence after mastectomy, recurrence in the conserved breast occurs at a more constant rate (approximately <1% per year), and is much less commonly associated with distant relapse. It is usually detected clinically and/or mammographically. Further surgery (most commonly mastectomy) is recommended and, after this, the patient may have no further recurrence.

Metastatic disease

Approximately 8% of patients have distant metastases at first presentation, but subsequently most patients with distant metastases will have been previously treated for breast cancer that was apparently localized to the breast and the axillary lymph nodes some time previously.

Patients with metastatic breast cancer are incurable. Treatment is therefore palliative, the aim being to relieve symptoms and maintain the best quality of life. Median survival after the diagnosis of visceral metastatic breast cancer is about 2 years, but there is a 'tail' of patients who survive for 5–10 years. Thus, management varies from the treatment of a patient with a rapidly fatal cancer to the management of what is a chronic, albeit serious, disease.

For the patient with metastatic disease, proper working of the multidisciplinary team is crucial. Her needs will vary over time and she will need access to a wide variety of disciplines. Most patients will be primarily under the care of their family doctor and the practice team, and the secondary sector should aim to support the GP so that the patient has prompt access to the surgeon, oncologist, orthopaedic specialist and palliative care team at the appropriate times.

The presentations of patients with metastases are legion, but among the commonest sites are bone (with pain or fracture; Figure 17.25), pleura (effusion causing dyspnoea; Figure 17.26), liver (lethargy, nausea, anorexia), peritoneal cavity (ascites causing abdominal distension; Figure 17.27), lung (dyspnoea, dry cough), supraclavicular fossa/cervical nodes (Figure 17.24),

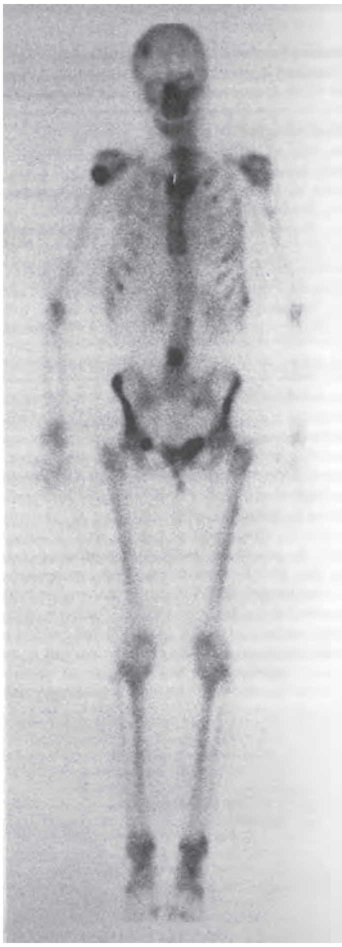


Figure 17.25 Staging bone scan showing metastases in the right humerus, thoracic and lumbar spine, pelvis and skull.

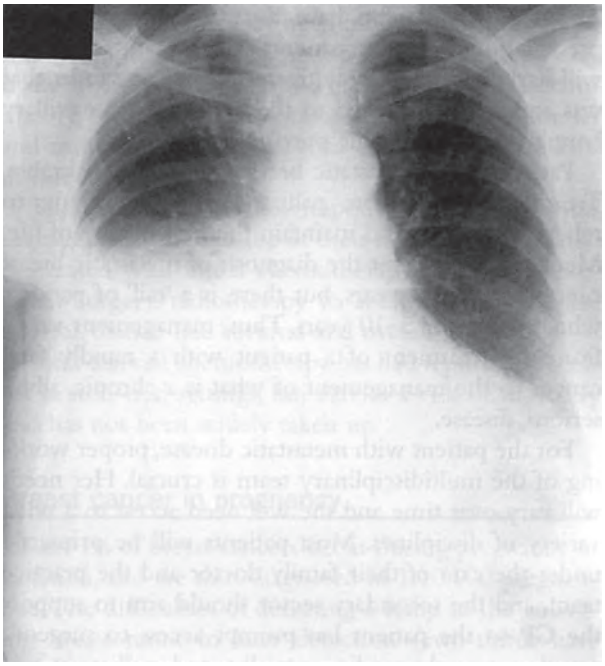


Figure 17.26 Pleural effusion from secondary breast cancer.

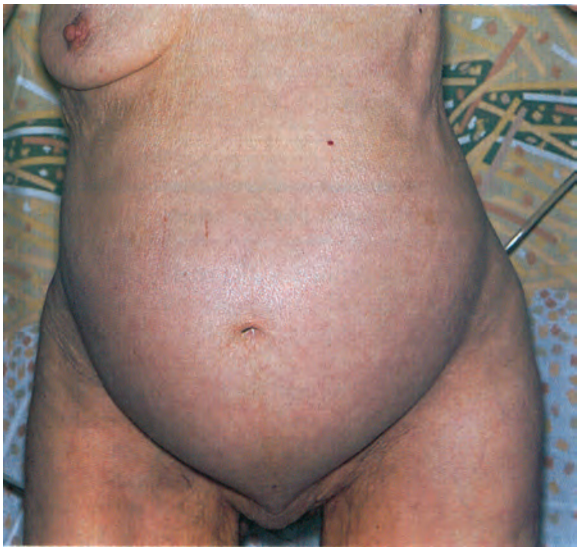


Figure 17.27 Gross ascites from metastatic breast cancer; previous left mastectomy.

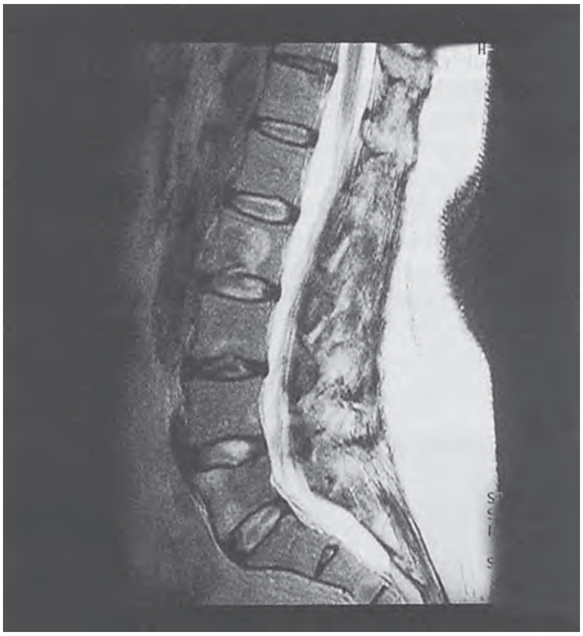


Figure 17.28 Spinal metastases (lumbar vertebra) on sagittal MRI scan.

spine (Figure 17.28), brain (headaches, imbalance, fits) and marrow (lethargy, infections, anaemia). Whatever the site of the initial metastasis, patients should be staged to assess the extent of metastatic spread by at least a full blood count, biochemistry (including bone and liver chemistry), CT of the chest, abdomen and pelvis; tumour markers (CEA, CA15-3) can be used to monitor the disease.

In general, patients presenting with visceral metastases (e.g. liver, lymphangitic carcinomatosis of lung) will have aggressive, usually endocrine-insensitive, disease with a poor prognosis, whereas relapse in soft tissue, pleura and/or bone is associated with a survival measured in many months or years (and often hormone-sensitive disease).

Principles of treatment

Since there is no evidence that treating asymptomatic metastases alters survival, treatment is aimed at symptom control and can be considered as:

- control of specific symptoms
- systemic therapy to control the disease
- local therapies for local symptoms.

Symptom control

Symptom control should be used in combination; pain control comprises analgesic regimens including non-steroidal anti-inflammatory drugs (NSAIDs), opioid drugs, amitriptyline/carbamazepine/gabapentin, transcutaneous electrical nerve stimulation and psychological control techniques.

Hypercalcaemia often presents as non-specific deterioration of health, confusion, abdominal symptoms, dehydration and ultimately renal failure and coma. Rehydration (oral, intravenous), bisphosphonates and therapy against the underlying cancer is required.

Systemic therapy

Since metastatic breast cancer is a systemic disease, the best way to control it is with effective systemic therapy.

In general, endocrine therapy is considered as first-line therapy, since it is less toxic than cytotoxic chemotherapy. The best predictor of response to endocrine therapy is the hormone receptor status of the primary tumour. A positive ER status is associated with response in about 60% of patients; response to second-line endocrine treatment is half that seen to first-line treatment. Thus, patients with ER-positive tumours will normally be treated with endocrine therapy in the first instance, unless they have advanced visceral metastases (e.g. liver), when the response to hormonal therapy is low and the delay in waiting for a hormonal response may compromise the patient's survival. Such patients should normally be considered for chemotherapy. Patients with ER-negative tumours should be considered for chemotherapy only since endocrine therapy is unlikely to help them; a balance must be struck between achieving response and limiting side effects. Response in ~50% patients (again with response to second-line treatments half of this) may achieve control for 6–10 months.

BOX 17.7 Standard treatments for breast cancer

- Premenopausal or perimenopausal women:
  - Women with low-risk disease should be considered for tamoxifen if ER positive. Aromatase inhibitors should be avoided in premenopausal patients
  - Women with intermediate- or high-risk disease who are ER positive should be offered chemotherapy and tamoxifen
  - Women with intermediate- or high-risk disease who are ER negative should be offered adjuvant chemotherapy including an anthracycline and/or a taxane. Tamoxifen or ovarian ablation holds no benefit
- Postmenopausal women:
  - Women with low-risk disease should be considered for tamoxifen or an aromatase inhibitor if ER positive
  - Women with intermediate- or high-risk disease and ER-positive tumours should be considered for chemotherapy followed by an aromatase inhibitor
  - Women with intermediate- or high-risk disease and ER-negative tumours should, if fit, be considered for chemotherapy

Endocrine therapy

In premenopausal patients with hormone receptor-positive metastatic breast cancer, tamoxifen displays clinical efficacy similar to ovarian ablation. Single agent tamoxifen has been associated with response rates around 30% and a median time to progression of 6 months in both pre- and postmenopausal patients with hormone receptor-positive metastatic breast cancer, with some patients experiencing prolonged stabilization of their disease.

Unlike tamoxifen and raloxifen, fulvestrant is considered a 'pure antioestrogen' because the drug downregulates the ER without agonist properties. Fulvestrant's efficacy has only been established in postmenopausal women with hormone receptor-positive metastatic breast cancer, in whom the drug has demonstrated similar efficacy to aromatase inhibitors. Additional studies have demonstrated improved efficacy with higher doses of fulvestrant.

In the metastatic setting, anastrozole demonstrated similar response rates to tamoxifen but prolonged time to progression when used as first-line therapy for the treatment of postmenopausal women with ER-positive metastatic breast cancer. Letrozole improved both response and time to progression when compared with tamoxifen as first-line treatment for hormone receptor-positive metastatic breast cancer. Similar results were seen with exemestane compared with tamoxifen and the risk of progression was also significantly reduced. Exemestane has also demonstrated activity in patients with metastatic breast cancer refractory to either anastrozole or letrozole, reflecting the different structure and mode of action of exemestane.

Cytotoxic chemotherapy

The principles of using cytotoxic chemotherapy for advanced breast cancer are similar to those for adjuvant therapy (page 400). The choice of cytotoxic agent will be determined by prior adjuvant therapy and the patient's general condition and fitness. The risk of anthracycline cardiotoxicity increases with cumulative dose, so, if anthracyclines have been used adjuvantly, there will, in general, be little scope to use them in the advanced disease setting. In general, first-line therapy for advanced disease patients will be an anthracycline-based combination (unless used adjuvantly) and second line would be a taxane or capecitabine. For patients with good general condition and chemosensitive disease, then third- and fourth-line therapy with drugs such as vinorelbine or gemcitabine can be useful.

Biological therapy

As a single agent, trastuzumab has demonstrated response rates of 15–34% in patients with metastatic HER2-overexpressing breast cancer. A pivotal trial in the metastatic setting demonstrated that the addition of trastuzumab to chemotherapy (either doxorubicin/epirubicin and cyclophosphamide or paclitaxel) improved response rate, median duration of response and overall survival. Because of a higher rate of congestive heart failure in patients treated with trastuzumab in combination with anthracyclines, paclitaxel became the initial chemotherapy backbone approved for use with trastuzumab. Since the initial approval, trastuzumab has been combined with other chemotherapy agents, such as docetaxel, navelbine, capecitabine, carboplatin and gemcitabine in the metastatic setting.

Approximately 60% of tumours that are HER2 positive are also hormone receptor positive. In metastatic breast cancer, the addition of trastuzumab improved response and progression-free survival when combined with anastrozole as first-line therapy in postmenopausal patients with hormone receptor-positive, HER2-positive metastatic breast cancer. The role for continued trastuzumab beyond disease progression remains poorly defined, but the existing limited data suggest that continuation of trastuzumab improves clinical outcome.

Lapatinib is a small molecule inhibitor of the kinase domain of both HER1 and HER2. Early phase studies demonstrated response for both single agent lapatinib and lapatinib given in combination with trastuzumab. A larger randomized study in women with HER2-positive advanced breast cancer who had disease progression while receiving trastuzumab demonstrated improved progression-free and overall survival for the combination of lapatinib and trastuzumab compared with lapatinib alone.

Lapatinib has been combined with endocrine therapy and found to improve clinical benefit and progression-free survival, compared with hormonal therapy alone, in patients with hormone receptor-positive, HER2-positive metastatic breast cancer. Combination therapy with capecitabine and lapatinib improved response (22% versus 14%) and time to disease progression (8.4 versus 4.4 months, compared with single agent capecitabine) in patients with HER2-positive metastatic breast cancer who had received prior anthracyclines, taxanes and trastuzumab. A large randomized phase III trial also evaluated the efficacy of paclitaxel with/without lapatinib in patients with HER2-positive or HER2-uncharacterized metastatic breast cancer. Although there were no significant differences in response, time to progression or overall survival in the intent to treat population, a preplanned, retrospective evaluation of HER2 status identified 86 (15%) patients with HER2-positive tumours. In this small subset of patients, response and time to progression were significantly improved with the addition of lapatinib to paclitaxel. Randomized trials investigating the role of lapatinib in the adjuvant setting are currently accruing patients.

Single agent lapatinib is well tolerated with manageable toxicities of skin rash and diarrhoea, probably related to the drug's inhibition of EGFR. Lapatinib also does not seem to share the cardiac toxicity of trastuzumab with cardiac event rates (symptomatic or asymptomatic decreases in left ventricular ejection fraction) of 2.2% and 1.7% in patients who had received prior anthracycline and trastuzumab, respectively.

Bisphosphonates

Bone is a common site of breast cancer metastasis and skeletal complications can reduce quality of life. Up to 70% of breast cancer patients with bone metastasis experience at least one skeleton-related event, with approximately 50% experiencing a pathological fracture. Metastatic cells stimulate osteoclast-mediated bone resorption and growth factors released from resorbed bone promote tumour growth. Bisphosphonates inhibit osteoclast-mediated bone resorption and are considered standard of care for the treatment of cancer-induced hypercalcaemia. Treatment with bisphosphonates reduces the risk of skeleton-related events by ~17% in patients with clinically evident bone metastasis.

Additionally, treatment of women with advanced breast cancer and no evidence of bone metastasis with oral clodronate has been demonstrated to reduce bone metastasis and other skeletal events, such as fracture. Adjuvant therapy with clodronate has demonstrated mixed results with two trials demonstrating a significant reduction (~30%) in the risk of developing bone metastasis in patients with operable localized breast cancer and a second study demonstrating no benefit in risk reduction of skeletal metastasis, but an improvement in overall survival in favour of clodronate in the adjuvant setting. Larger trials investigating the effects of bisphosphonates in the adjuvant setting are required to give definitive evidence for their use in reducing bone metastasis and/or osteoporosis from prolonged use of aromatase inhibitors. In general, bisphosphonate therapy is well tolerated, with some patients reporting bone pain or flu-like symptoms with initial administration; however, serious complications such as renal failure and osteonecrosis of the jaw have been reported in a small subset of patients receiving therapy with these agents.

Local therapies

Patients need careful assessment of their symptoms and many may be helped by appropriate local surgical interventions. Examples include:

- Drainage of pleural effusions (Figure 17.26) plus instillation of talc as a tumoricidal and pleuradhesis agent.
- Drainage of ascites (Figure 17.27).
- Surgical stabilization of sites of bone weakness due to metastases – most commonly in the long bones. Prophylactic fixation is preferred, but some patients will need fixation following pathological fracture.
- Palliative radiotherapy, most commonly for painful bony metastases (Figure 17.28). A single fraction of 7–8 Gy will give substantial pain relief in 80% of patients but will not be effective if there is mechanical pain. NSAIDs may confer additional benefit.
- Spinal cord compression. This is an oncological emergency and patients require rapid neurosurgical/radiotherapy assessment, usually following an urgent MRI scan to show the nature and level of the block(s) within the spinal canal.
- Brain metastases. High-dose corticosteroids (4 mg q.i.d. dexamethasone) are usually employed to reduce oedema followed by radiotherapy. Local excision plus radiotherapy can be effective for single metastases depending on the site.

- Surgical resection of metastases. A single brain metastasis in a patient who is otherwise reasonably fit should be considered for resection (followed by palliative radiotherapy) in order to optimize local control. Very occasionally, isolated other metastases (e.g. in liver, bone or lung) should be considered for resection but this is an unusual clinical scenario.
- Surgical resection of the primary tumour. In patients with limited extent of metastatic disease that is responding to systemic therapy, resection of the primary tumour should be considered to obtain local control. There is some evidence from retrospective series (with the caution that this implies) that this may improve survival.

Male breast cancer

The incidence of male breast cancer increases with age but still accounts for <1% of breast cancers, and in the USA is the cause of death in only one in 1 000 000 men.

Risk factors include Klinefelter syndrome (testicular atrophy and insufficiency, gynaecomastia), in which the ratio of oestrogen to androgen is elevated and increases the risk of breast cancer some 20 times. *BRCA2* mutation in breast cancer families is associated with breast cancer in male members of the families and hence molecular screening of male relatives in such families may be justified.

Presentation is as a palpable, painless, firm lump that is often eccentric to the nipple and fixed to the chest wall and skin (Figure 17.29). There may be erythema progressing to ulceration through the skin, nipple retraction or a bloody nipple discharge.

Diagnosis

As for female breast cancer, diagnosis is by triple assessment: clinical examination, mammography and core biopsy to allow the distinction between gynaecomastia and male breast cancer. Staging is required if T3 or above.

Histologically, most male breast cancer is ductal; DCIS accounts for 10%; lobular cancer and special types are comparatively rare. Almost all male breast cancers are ER positive, but should be tested by immunohistochemistry.

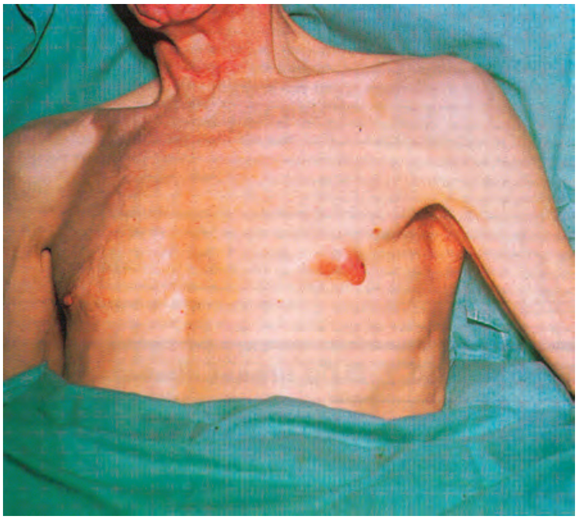


Figure 17.29 Male breast cancer (T4 disease).

Treatment

Modified radical mastectomy, or radical mastectomy if the pectoralis major muscle is directly infiltrated by breast cancer, accompanied by axillary node clearance is the standard treatment even for advanced disease involving the skin or underlying chest wall muscle. Radiotherapy is based on data from female breast cancer, and is used to assist local disease control together with surgery.

While the use of adjuvant endocrine or chemotherapy has not been the subject of the major clinical trials undertaken for female breast cancer, 65–85% of male breast cancers are ER positive and 67% PgR positive. Not surprisingly, endocrine therapy can be effective: tamoxifen (with a 70% response rate) or letrozole and, as second-line therapy, cyproterone acetate or luteinizing hormone releasing hormone analogues have largely replaced oestrogens as therapy. Anthracycline-based chemotherapy, as for female breast cancer, is used in both the adjuvant and palliative settings.

Follow-up

Most patients with breast cancer will complete their primary treatment (surgery, radiotherapy, chemotherapy, etc.) and return to the community. Traditionally, they have then been followed up in the hospital outpatient department. The aim of follow-up is to detect local recurrence (when treatable) or contralateral disease. For those patients with established metastatic disease and/or uncontrolled local disease, continued follow-up is appropriate since they are likely to need ongoing care tailored to the pace of the disease and their needs. For those patients who are apparently disease free, the purpose of follow-up is less clear, but includes detection of local disease recurrence or detection of a second, contralateral cancer or metastatic disease, assessment of treatment morbidity or simply providing reassurance.

Some patients find regular follow-up reassuring but others find it stressful. The optimal frequency and length of follow-up is not defined. It is, however, important that there is a strategy of care agreed between the patient, the GP and the hospital, so that the appropriate care can be offered to the patient should any problems develop.

Detection of distant metastases

Metastatic relapse will present unpredictably and usually between clinic visits. Routine screening for metastases in patients who are asymptomatic does not improve survival. Thus, investigations for metastases should be restricted to patients with symptoms suggestive of metastatic relapse.

Detection of local recurrence

Relapse in the chest wall after mastectomy or in the axilla or supraclavicular fossa is usually detected clinically. Local recurrence after conservation occurs at a constant rate each year, hence follow-up should be continued to 10 years. Following mastectomy the risk of local recurrence is greatest in the first 2–3 years and decreases thereafter. Follow-up may be discontinued, if disease free, at 5 years. Although these relapses may be found at a routine hospital visit, they may also be found by the patient between visits. It is important that the patient can be referred back promptly to the clinic.

Relapse in the treated breast (following conservation treatment) may be found clinically or (more commonly) on follow-up mammography. The latter is therefore recommended every 1–2 years.

Relapse in the contralateral breast

Breast cancer in one breast is associated with an ~1% per annum increased risk of developing cancer in the contralateral breast. The other breast should therefore be examined at each visit and a mammogram performed every 1–2 years.

Morbidity of therapy

Follow-up is also an opportunity to assess the morbidity of treatment, especially locoregional or systemic effects.

Specific considerations include the following.

Arm mobility

Women with breast cancer may develop arm stiffness directly related to their surgery and radiotherapy. Women undergoing breast and/or axillary surgery need shoulder exercises to enable them to recover a full range of arm and shoulder mobility.

Lymphoedema

Lymphoedema of the upper limb may occur in women with breast cancer secondary to lymphatic damage caused by surgery and/or radiotherapy, or because of obstruction caused by local tumour (Figure 17.30). Thus, all patients undergoing surgery and/or radiotherapy treatment to the axilla should receive pretreatment information on lymphoedema. The incidence of lymphoedema has been cited at between 5% and 60%, depending on treatment combinations. Although there is currently no agreed surgical procedure to cure lymphoedema, it is possible to reduce the size of the arm. The most effective management and maintenance comprises multimodal physical therapy (skin care, external support, exercise, massage) and education. Diuretics or pneumatic compression pumps should not be relied upon. Lymphoedema should be treated at the first sign of swelling when management will be more effective. The lymphoedematous arm is prone to streptococcal infection after minor injury (e.g. gardening); patients should be instructed to seek medical attention, including antibiotics, at the first signs of infection (tenderness, redness, increased swelling, pain on movement of the limb).

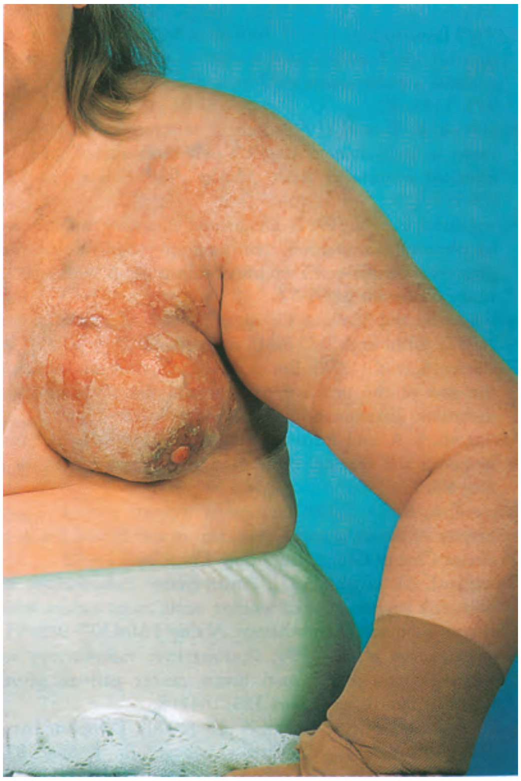


Figure 17.30 Lymphoedema of the left upper limb secondary to advanced breast cancer with chest wall disease and infiltration of the axilla. Note the armlet to symptomatically treat the lymphoedema.

Menopausal symptoms

Menopausal symptoms are a recognized issue for women with breast cancer. Many women who receive treatment for breast cancer subsequently experience menopausal symptoms, either as a result of their adjuvant treatment or as a natural process. The average age at menopause is 50 years, but in 25% of women who develop breast cancer premenopausally and undergo adjuvant treatment it is 10 years earlier. Although HRT is widely advocated for the treatment of menopausal symptoms, its use in women with a personal or family history of breast cancer remains controversial, and alternative methods of coping with the menopause have not yet been fully explored. Some studies show that progestogens such as megestrol acetate and soya protein are useful in alleviating menopausal symptoms.

Women should be informed regarding the potential effect of cancer treatment on their menopausal status, and advised regarding non-oestrogen alternatives and self-care strategies that might alleviate their symptoms.

Prostheses

External breast prostheses are the most common method of restoring breast symmetry following surgery, and a wide variety of shapes and sizes are available. A soft temporary breast prosthesis should be fitted before hospital discharge, and a permanent prosthesis should be fitted either 6–8 weeks postoperatively (or when the wound is fully healed) or following completion of radiotherapy once any skin reaction has resolved. Women should also be given advice on bras, swimwear and replacement prostheses.

Psychological considerations

Women (and men) with breast symptoms may suffer considerable emotional distress, even once a benign diagnosis is made.

Breast cancer is a high-profile disease, the progression and treatment of which results in substantial physical and psychological morbidity.

To participate effectively in decision-making, women and their families need to receive adequate information and sufficient time to discuss treatment options. Breast care nurses and voluntary sector organizations can provide verbal, written and multimedia information. Studies of the physical impact of screening and genetic counselling appear to be reassuring. There is no clear evidence that mastectomy and breast conservation differ in terms of clinically significant anxiety or depression; some studies have found that breast conservation is associated with fewer problems related to body image, although conservation may carry an increased risk of worrying about recurrence. Hence, some women choose mastectomy, although not all women wish to share in the decision-making. Significant numbers of women develop psychological problems after anticancer therapy treatment and some require formal treatment. Psychological factors may influence the course of malignant disease but can certainly be harnessed to combat the side effects of treatment, particularly chemotherapy. Disease recurrence is associated with increased psychological morbidity; formal psychological interventions can be used successfully should it occur.

Acknowledgements

The authors thank Mr J.H. Stevenson (Figures 17.16, 17.17, 17.18), Dr C. Purdie (Figures 17.2c, 17.9, 17.10, 17.11, 17.12), Mr D.C. Brown (Figure 17.21), Dr D. McLean (Figures 17.3, 17.19) and Dr Sarah Vinnicombe (Figures 17.13, 17.23, 17.24 and 17.27) for permission to use clinical slides.

GUIDE TO FURTHER READING

Alvarez S, Añorbe E, Alcorta P, *et al.* Role of sonography in the diagnosis of axillary lymph node metastases in breast cancer: a systematic review. *AJR Am J Roentgenol* 2006;**186**:1342–1348.

Association of Breast Surgery at BASO. Surgical guidelines for the management of breast cancer. *Eur J Surg Oncol* 2009;**35**:s1–s22.

Bartelink H, Horiot J-C, Poortmans P, *et al.* Recurrence rates after treatment of breast cancer with standard radiotherapy with or without additional radiation. *N Engl J Med* 2001;**345**:1378–1387.

Bates T, Riley DL, Houghton J, *et al.* Breast cancer in elderly women: a Cancer Research Campaign trial comparing treatment with tamoxifen and optimal surgery with tamoxifen alone. *Br J Surg* 1991;**78**:591–594.

Beatson GT. On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment with illustrative cases. *Lancet* 1896;**ll**:104–107.

Berry DA, Cronin KA, Plevritis SK, *et al.* Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;**353**:1784–1792.

Blamey RW. The design and clinical use of Nottingham Prognostic Index in breast cancer. *Breast* 1996;**5**:156–157.

Dixon JM, Thompson AM. Effective surgical treatment for mammary duct fistula. *Br J Surg* 1991;**78**:1185–1186.

Early Breast Cancer Trialists' Collaborative

Early Breast Cancer Trialists' Collaborative Group. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;**366**:2087–2106.

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer? *Genet Med* 2009;**11**:66–73.

Fisher B, Dignam J, Wolmark N, *et al.* Tamoxifen in treatment of intraductal breast cancer: National Surgical Adjuvant Breast and Bowel Project B-24 randomised controlled trial. *Lancet* 1999;**353**:1993–2000.

Fisher B, Costantino JP, Wickerham DL, *et al.* Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst* 2005;**97**:1652–1662.

Hermansen C, Skovgaard Poulsen H, Jensen J, *et al.* Diagnostic reliability of combined physical examination, mammography and fine-needle puncture ('triple-test') in breast tumours. A prospective study. *Cancer* 1987;**60**:1866–1871.

Howell A, Cuzick J, Baum M, *et al.* Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet* 2005;**365**:60–62.

Koscielny S, Tubiana M, Lê MG, *et al.* Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination. *Br J Cancer* 1984;**49**:709–715.

Kuhl C, Weigel S, Schrading S, *et al.* Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer: the EVA trial. *J Clin Oncol* 2010;**28**:1450–1457

Mandelblatt JS, Cronin A, Bailey S, *et al.* Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med* 2009;**151**: 738–747.

Mansel RE, Fallowfield L, Kissin M, *et al.* Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC Trial. *J Natl Cancer Inst* 2006;**98**: 599–609.

McIntosh SA, Purushotham AD. Lymphatic mapping and sentinel node biopsy in breast cancer. *Br J Surg* 1998;**85**:1347–1356.

Moss SM, Cuckle H, Evans A, *et al.* Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* 2006;**368**:2053–2060.

O'Flynn EA, Wilson AR, Michell MJ. Image-guided breast biopsy: state-of-the-art. *Clin Radiol* 2010;**65**:259–270.

Overgaard M, Hansen PS, Overgaard J, *et al.* Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. *N Engl J Med* 1997;**337**:949–955.

Overgaard M, Jensen MB, Overgaard J, *et al.* Postoperative radiotherapy in high-risk post-menopausal breast cancer patients given adjuvant tamoxifen. *Lancet* 1999;**353**:1641–1648.

Pain SJ, Purushotham AD. Lymphoedema following surgery for breast cancer. *Br J Surg* 2000;**87**:1128–1141.

Piccart-Gebhart MJ, Procter M, Leyland-Jones B, *et al.* Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;**353**:1659–1672.

Rapiti E, Verkooijen HM, Vlastos G, *et al.* Complete excision of primary breast tumour improves survival of patients with metastatic breast cancer at diagnosis. *J Clin Oncol* 2006;**24**:2743–2749.

Romond EH, Perez EA, Bryant J, *et al.* Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;**353**:1673–1684.

Ross JR, Saunders Y, Edmonds PM, *et al.* Systematic review of role of bisphosphonates on skeletal morbidity in metastatic disease. *BMJ* 2003;**327**:469–472.

Rosselli Del Turco M, Palli D, Cariddi A, *et al.* Intensive diagnostic follow-up after treatment of primary breast cancer: a randomized trial. National Research Council Project on Breast Cancer follow up. *JAMA* 1994;**271**:1593–1597.

Slamon DJ, Leyland-Jones B, Shak S, *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;**344**:783–792.

Smith IE, Dowsett M, Ebbs SR, *et al.* Neoadjuvant treatment of postmenopausal breast cancer with anastrozole, tamoxifen, or both in combination: the Immediate Preoperative Anastrozole, Tamoxifen, or Combined with Tamoxifen (IMPACT) multicenter double-blind randomized trial. *J Clin Oncol* 2005;**23**:5108–5116.

Smith IE, Dowsett M, Yap YS, *et al.* Adjuvant aromatase inhibitors for early breast cancer after chemotherapy induced amenorrhoea: caution and suggested guidelines. *J Clin Oncol* 2006;**24**:2444–2447.

START Trialists' Group. The UK Standardization of Breast Radiotherapy (START) Trial B of radiotherapy hypofractionation for treatment of early breast cancer: a randomized trial. *Lancet* 2008;**371**:1098–1107.

Thompson AM. Axillary node clearance for breast cancer. *J R Coll Surg Edin* 1999;**44**:111–117.

Veronesi U, Galimberti V, Zurrida S, *et al.* Prognostic significance of number and level of axillary node metastases in breast cancer. *Breast* 1993;**2**:224–228.

Vogel CL, Cobleigh MA, Tripathy D, *et al.* Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;**20**:719–726.

Willett A, Michell M, Lee M (eds). *Best Practice Diagnostic Guidelines for Patients Presenting with Breast Symptoms.* London, UK: Department of Health, 2010. Available from: http://www.rcrbreastgroup.com.

Wishart GC, Azzato EM, Greenberg DC, *et al.* PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res* 2010;**12**:R1.

CHAPTER 18

Disorders of the thyroid gland

SABA P. BALASUBRAMANIAN AND DAVID SMITH

Introduction

Thyroid disease encompasses a wide variety of inflammatory, autoimmune, nutritional and neoplastic pathologies. Although the disease states are defined using well-established criteria, there is a usually a continuum of change within the spectrum ranging from a normal thyroid (structurally and physiologically) to the extremes of disease phenotypes. Understanding where an individual patient lies in this spectrum and initiating an optimal management plan requires a detailed knowledge of the basic sciences and pathophysiology of thyroid disease.

The epidemiology of thyroid disease is very variable. Hypothyroidism is the second commonest endocrine disease and thyroid nodules are widely prevalent in most populations. However, thyroid cancer is an uncommon malignancy and certain types of thyroid cancers [such as medullary thyroid cancer (MTC) and anaplastic thyroid cancer] are very rare. Although surgical management is only required in a small minority of patients with thyroid disease, thyroid surgeons should have an in-depth understanding of the management of patients with 'non-surgical' thyroid disease. We also recommend a multidisciplinary approach bringing together endocrinologists, radiologists, oncologists and surgeons to facilitate discussion and appropriate management of patients with complex benign disease and thyroid cancers.

This chapter starts with a discussion of thyroid embryology, anatomy and physiology followed by sections on clinical problems in thyroid disease – hypothyroidism, hyperthyroidism, thyroid nodules and thyroid cancer.

Thyroid embryology, anatomy and physiology

Thyroid embryology

The thyroid is principally of *endodermal* origin, and is derived from the floor of the pharynx between the tuberculum impar (the medial swelling of the tongue) and the cupola. An endodermal track migrates down the neck in front of the primitive larynx and divides into right and left branches, which develop into the lateral lobes of the thyroid gland. This track remains connected to the tongue (at a point called the foramen caecum, located at the junction of the anterior and posterior parts of the tongue) by the thyroglossal duct. The duct normally closes at around the fifth week, and the only remnant is a shallow blind pit at the foramen caecum. Remnants of the duct may, however, persist and develop into well-differentiated thyroid tissue (ectopic thyroid) or a cystic lesion (thyroglossal cyst). Occasionally, these may be associated with absence of the thyroid in its orthotopic position, reflecting arrest of migration of the thyroglossal duct. Complete failure of descent may result in a lingual thyroid, at the back of the tongue. Papillary carcinoma of the thyroid may rarely develop in an ectopic thyroid gland.

The thyroid is in its normal position just inferior to the cricoid cartilage by the seventh week. Follicles appear and the thyroid begins to secrete hormones by the twelfth week, one of the first organs to do so. The thyroid also has a *neural crest* origin as cells from the neural crest migrate into the ultimobranchial body, which is derived from the ventral part of the fourth/fifth pharyngeal pouch. The ultimobranchial body becomes incorporated into the thyroid and has been thought to develop into the tubercle of Zuckerkandl on the lateral aspects of the thyroid lobes. However, another view is that this may not contribute to any thyroid tissue. The 'neural crest' cells in the ultimobranchial body develop into the parafollicular cells, which secrete calcitonin. Hence, the thyroid gland is derived from both the endoderm of the primitive pharynx (follicular cells) and the ectoderm, i.e. the neural crest (parafollicular cells).

Thyroid anatomy

The thyroid is a highly vascular gland, weighing about 15 g, and consists of two lobes, united in the midline by the isthmus, which overlies the second and third tracheal rings. There may also be a pyramidal lobe, superior to the isthmus and often to the left of the median plane. This is present in 50% of individuals and is the remnant of the thyroglossal tract. A fibrous capsule surrounds the thyroid and extends into the gland as septae. This capsule is enveloped by the visceral layer of the pretracheal cervical fascia. Posteriorly, the gland is attached to the cricoid cartilage and the superior tracheal rings by dense connective tissue (Berry's ligament). This attachment of the thyroid to the trachea causes the thyroid to move on swallowing, which helps distinguish thyroid nodules from other neck lumps.

The arterial supply is from the superior and inferior thyroid arteries, which lie in the plane between the thyroid capsule and the pretracheal fascia. The superior thyroid artery originates from the external carotid artery and divides into an anterior and posterior branch at the superior pole of the thyroid. The external branch of the superior laryngeal nerve (EBSLN) runs inferolateral to the oblique line of thyroid cartilage and has a close relationship with the superior thyroid artery. It typically runs over the inferior constrictor muscle to reach and supply the cricothyroid (in some cases, the nerve passes deep to inferior constrictor muscle and is therefore not routinely seen at operation). Care must be taken to ligate the superior thyroid artery on the capsule of the thyroid at operation and to be aware of the possibility of damage to the EBSLN when applying diathermy to the small branches of the superior thyroid artery. Damage to the EBSLN results in paralysis of the cricothyroid muscle. This muscle tenses the vocal cords, which is important in production of high pitch, as in singing. The inferior thyroid artery originates from the thyrocervical trunk, which arises from the first part of the subclavian artery. It enters the thyroid on its lateral aspect and has a close relationship with the recurrent laryngeal nerve (RLN). On the left, the RLN runs vertically along the tracheo-oesophageal groove and is usually medial to the inferior thyroid artery. On the right, the RLN runs a more lateromedial course and has a more intimate relationship to the branches of the inferior thyroid artery. Superiorly, the nerve passes through the posterior portion of Berry's ligament (an area where it is prone to damage) before entering the larynx, usually adjacent to the inferior cornu of the thyroid cartilage. The RLN is the motor supply to the intrinsic muscles of the larynx. An injury to the nerve causes ipsilateral vocal cord palsy. In about 1% of cases there is a non-recurrent right laryngeal nerve, owing to a vascular anomaly in the development of the aortic arches. In 10% of patients there is a thyroidea ima artery, which arises from the brachiocephalic trunk, and passes into the inferior border of the isthmus.

The venous drainage of the thyroid is via the superior and middle thyroid veins, which drain into the internal jugular and the inferior thyroid veins, which drain into the brachiocephalic vein. The lymphatic drainage of the thyroid is to the prelaryngeal, pretracheal and paratracheal lymph nodes medially and then to the deep cervical lymph nodes laterally around the carotid sheath.

Thyroid physiology

The thyroid gland consists of two main types of cells – follicular and parafollicular cells. The follicle is the basic functional unit of the thyroid and consists of a single layer of cuboidal (follicular) cells around a store of colloid. The follicular cells synthesize thyroid hormone in four stages, which include:

- *Iodide trapping.* Iodide is actively transported (by an ATP-dependent process) into the thyroid follicular cell and then to the apical membrane. The average daily requirement of iodine is 150 g. The normal daily Western dietary intake contains approximately 500 g.
- *Organification.* The iodide is oxidized by the enzyme thyroid peroxidase and then combined with tyrosine to form the inactive iodotyrosines: 3-monoiodotyrosine (MIT) and 3,5-di-iodotyrosine (DIT). The iodotyrosines are incorporated into the soluble protein, thyroglobulin, which is stored as colloid in the follicular lumen of the thyroid.
- *Coupling.* The iodotyrosines in the thyroglobulin are then coupled to form triiodothyronine or T_3 (MIT and DIT) and thyroxine or T_4 (two molecules of DIT).
- *Release.* Colloid is taken up by the thyroid cell by endocytosis to form endosomes. The thyroglobulin is then hydrolysed to liberate T_4, T_3, MIT and DIT. The MIT and DIT are deiodinated and the released iodide reused by the thyroid cell. The active hormones, T_4 and T_3, are secreted into the blood.

The vast majority of the released thyroid hormone is in the form of T_4 (90%). Most of the T_3 (80–90%) is produced by the peripheral conversion of T_4. T_3 is much more potent than T_4. The metabolic activity of thyroid hormone is determined by the amount of free T_3 and free T_4. Thyroxine is very highly protein bound in plasma (99.95% bound to thyroid-binding globulin, transthyretin and albumin, with about 0.05% free). When bound, T_4 is not physiologically active but provides a storage pool of thyroid hormone, which can last 2–3 months (mean half-life of T_4 is 6.5 days). Reverse T_3 (rT_3) is also produced by the deiodination of T_4. It is not physiologically active and increased levels of rT_3 are produced in hyperthyroidism, and periods of excess catabolism (e.g. burns, sepsis).

Peripheral action of thyroid hormone

The principal effects of thyroid hormones are to facilitate normal growth and differentiation and to increase the rate of metabolism. Thyroid hormones act predominantly via a nuclear thyroid receptor (TR), which upregulates gene transcription and thereby increases protein synthesis. T_4 is relatively inactive in the periphery, owing to a low affinity for TR in comparison with T_3.

Thyroid hormone regulation

Thyroid-stimulating hormone (TSH) is the major regulator of thyroid activity, with increased levels causing hypertrophy of the thyroid. TSH is secreted by the anterior pituitary, and is a glycoprotein with an and subunit (the subunit being common to follicle-stimulating hormone, luteinizing hormone and human chorionic gonadotrophin). TSH acts by binding to the TSH receptor on the follicular cell membrane, leading to increased thyroid hormone synthesis, predominantly via cAMP as the second messenger. Thyroid-releasing hormone (TRH) is the most important positive stimulus for the production of TSH. TRH is produced in the paraventricular nucleus of the hypothalamus and passes through the median eminence to the anterior pituitary via the hypophyseal portal system. T_3 has a negative feedback on both the anterior pituitary and the hypothalamus.

Parafollicular cells

The parafollicular, or C cells, secrete calcitonin. Calcitonin is a 32 amino acid peptide that lowers calcium largely by the inhibition of osteoclasts. It is of little physiological importance in humans as there is no disturbance of calcium regulation following thyroidectomy, provided that the parathyroids are preserved. It is a sensitive tumour cell marker for medullary thyroid carcinoma (see Differentiated thyroid carcinoma). Calcitonin also has a role in the treatment of Paget disease of the bone.

Hypothyroidism

Epidemiology

Hypothyroidism is one of the most common endocrine disorders, with approximately 5% of the female population developing hypothyroidism at some stage in their lives. The vast majority are caused by primary hypothyroidism, with secondary (anterior pituitary gland disorders) and tertiary (hypothalamic disorders) hypothyroidism being rare. Primary hypothyroidism is usually caused by autoimmune thyroiditis (the commonest among them being Hashimoto's thyroiditis), or by treatment of other thyroid disease [e.g. following surgery or radioactive iodine (RAI) for thyrotoxicosis]. About one-third of patients undergoing a hemithyroidectomy and up to 70% of patients treated by subtotal thyroidectomy are hypothyroid at 10 years, this being influenced by remnant size and the degree of lymphocytic infiltration of the thyroid. The use of RAI in the management of thyrotoxicosis results in 50% of patients developing hypothyroidism in the first year, with 3% per annum subsequent to that. Lifelong follow-up and monitoring of TSH levels are therefore required for these patients. External beam radiotherapy to the neck can result in subclinical or clinical hypothyroidism in up to 50% of patients (usually 2–7 years after treatment). In addition, certain drugs can cause hypothyroidism, including amiodarone and lithium.

Thyroiditis

Several different types of thyroiditis exist. Although there are overlapping features among the different types, they are commonly classified on the basis of histology and/or aetiology. These include:

- chronic lymphocytic (Hashimoto's thyroiditis)
- Graves' disease (see Thyrotoxicosis)
- subacute granulomatous thyroiditis
- subacute lymphocytic thyroiditis
- postpartum thyroiditis
- postsurgical thyroiditis (e.g. following parathyroid surgery)
- radiation thyroiditis (following RAI and external radiation)
- suppurative thyroiditis (or thyroid abscess)
- Reidl's thyroiditis.

The commonest form of thyroiditis and hypothyroidism in the population is a specific autoimmune form of thyroiditis called Hashimoto's thyroiditis. This has a familial predisposition. Females are more commonly affected (male to female ratio 9:1) with the usual age of onset being 40–50, although it can occur at any age. The usual clinical presentation is with painless thyroid enlargement (a firm, rubbery gland); occasionally the patient may present with pain or pressure in the neck. The gland enlarges owing to the lymphocytic infiltrates and areas of focal hyperplasia, caused by TSH stimulation. In the early stages, free T_4 and T_3 may be normal, with an elevated TSH (subclinical hypothyroidism). The patient may also present with thyrotoxicosis (4%), caused by release of thyroxine, before eventually becoming hypothyroid.

Clinical features

The development of clinical features is variable and often slow and insidious. A significant proportion of patients have mild or non-specific features and the diagnosis is made on biochemistry as part of screening for a variety of medical problems. Common symptoms of hypothyroidism include dry skin, coarse hair, hair loss, brittle nails, weight gain, hoarseness (unrelated to RLN palsy), cold intolerance, paraesthesia, muscle stiffness and aches, drowsiness, lethargy, easy fatigability and constipation. Signs include cool and dry skin, bradycardia, puffy face and hands and slow deep tendon reflexes. Clinical features are less obvious and often atypical in the elderly. In myxoedema (severe and life-threatening hypothyroidism), there is progressive deterioration in mental status, localized neurological signs, hypothermia, hypoglycaemia and hyponatraemia.

Subclinical hypothyroidism is a biochemical entity defined as raised TSH in the presence of normal levels of T_3 and T_4. These patients are usually asymptomatic.

Investigations

Investigations are largely biochemical. A thyroid profile usually includes TSH and free T_4 (free T_3 is sometimes done, especially if T_4 levels are normal). Antimicrosomal antibodies (antithyroid peroxidase) are found in 90% of patients with Hashimoto's thyroiditis, but are not specific as they are also found in 70% of patients with Graves' disease. Antithyroglobulin antibodies are found in 60% of patients and are more specific to Hashimoto's thyroiditis. Thyroid ultrasound or isotope scans are usually not required for the diagnosis of hypothyroidism.

Treatment

The aim of treatment is to normalize the TSH level by giving supplemental thyroxine. Thyroxine is absorbed in the small intestine and has a half-life of 7 days. Care must be taken when instituting therapy in the elderly or patients with coronary artery disease, when a lower dose should be initially administered (starting at 25 g and increasing by 25 g increments every 2 weeks). After TSH levels are stabilized, therapy is monitored by checking the TSH level on an annual basis.

The treatment of subclinical hypothyroidism is debatable and best done in conjunction with a medical endocrinologist. Consensus statements suggest thyroxine treatment in patients with TSH levels persistently above 10.0 mU/L.

Long-term consequences

Patients with autoimmune thyroiditis have an increased predilection for other autoimmune disorders (such as pernicious anaemia). Patients with Hashimoto's thyroiditis also have a higher risk of developing a thyroid lymphoma than the general population. Eighty per cent of all -cell thyroid lymphomas develop on a background of Hashimoto's thyroiditis (Figure 18.1).

Hyperthyroidism

Epidemiology

Hyperthyroidism refers to the state of increased synthesis and secretion of thyroid hormones. Thyrotoxicosis is defined as excess levels of circulating thyroid hormones, which includes hyperthyroidism but also other uncommon conditions in which high levels may be due to destructive thyroiditis and excessive release, ectopic secretion (from struma ovarii or choriocarcinoma) or excessive ingestion of thyroid hormones. Thyrotoxicosis is caused in the vast majority of instances by primary thyroid pathology (including Graves' disease, toxic multinodular goitre and solitary toxic adenoma).

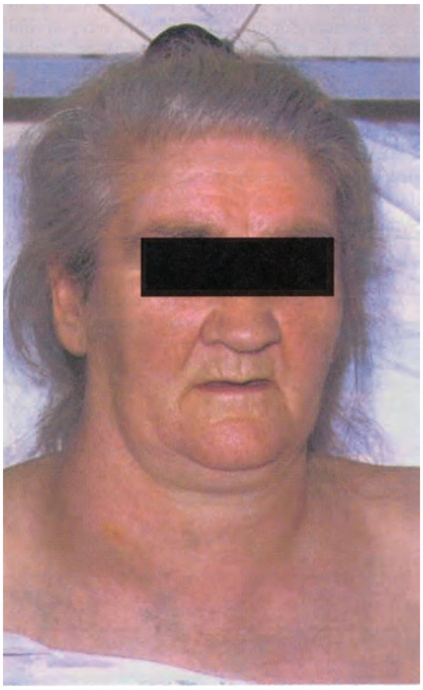


Figure 18.1 Patient who developed lymphoma on a background of Hashimoto's thyroiditis.

Graves' disease

Graves' disease (also called Basedow disease in mainland Europe) is an autoimmune disorder, characterized, in 90% of cases, by the presence of an immunoglobulin G antibody against the TSH receptor. The antibody is termed thyroid-stimulating immunoglobulin or TSH receptor antibody (TRAb) and stimulates the follicular cells, leading to excess production of thyroxine, hyperplasia and hypertrophy of the gland. Graves' disease is commoner in females (female to male ratio 7:1) and is not premalignant. The thyroid in patients with Graves' disease is usually diffusely enlarged, although sometimes multinodularity (or asymmetric enlargement) is noted. Thyroid cancer has occasionally been detected in surgically resected thyroid gland specimens of patients with Graves' disease. The nature of this association is uncertain and a previous presumption of increased aggressiveness of thyroid cancer in such patients has not been established.

Other causes of primary thyrotoxicosis

Plummer disease (described originally in 1913) is the term given to a hyperfunctioning thyroid, which may be either multinodular or a solitary nodule. A *toxic multinodular goitre* usually develops in a large, longstanding multinodular goitre of at least 10 years' duration. A *solitary toxic adenoma* is an autonomous nodule that produces enough thyroid hormone to cause hyperthyroidism. An unusual cause of thyrotoxicosis is the Jod–Basedow phenomenon, which is caused by the excessive release of thyroxine in an iodine-deficient patient on resumption of dietary iodine intake or administration of intravenous contrast. The phenomenon is most commonly observed in patients over 50 years with a longstanding multinodular goitre.

Clinical features

Symptoms include anxiety, tremor, palpitations, hyperactivity, increased appetite, weight loss, fatigue, heat intolerance, diarrhoea, menstrual irregularity and skin changes. Signs include warm and moist skin, irritability, restlessness and other psychiatric manifestations and features of a hyperdynamic circulation (and atrial fibrillation in the elderly). Other uncommon features include myopathy (Figure 18.2), gynaecomastia, decreased fertility, hyperpigmentation, splenomegaly and lymphadenopathy. Patients with Graves' disease may also have an infiltrative dermopathy often affecting the legs. This has been paradoxically termed 'pretibial myxoedema'.

Thyroid eye disease, which is often associated with Graves' disease (see below), manifests in its early stages as prominence of the eyes, irritation and increased watering of the eyes and

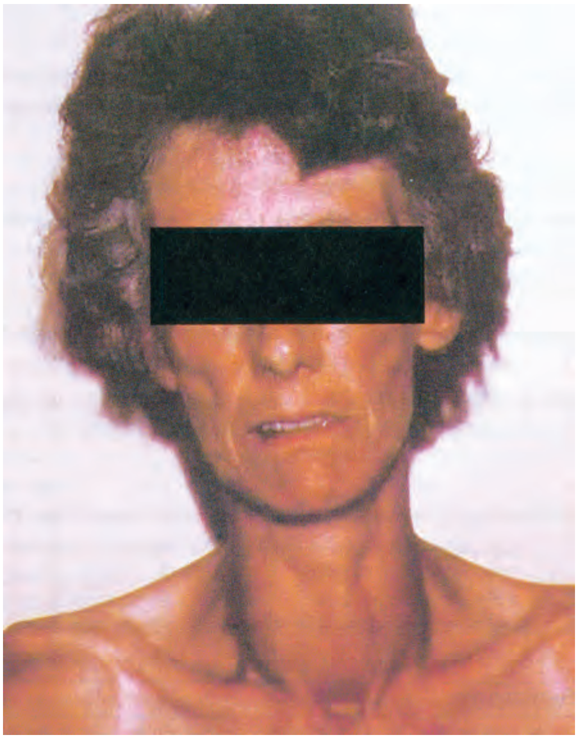


Figure 18.2 Proximal upper limb myopathy in a patient with thyrotoxicosis.

in later stages as periorbital oedema, chemosis (conjunctival oedema), lid and/or globe lag, ophthalmoplegia and visual loss.

Investigations

TSH levels are suppressed in thyrotoxicosis. In Graves' disease, antimicrosomal (thyroid peroxidase antibody) and antithyroglobulin antibodies are raised in 70% of patients, whereas the TRAb is raised in 90% of patients. Antibodies are usually absent in toxic multinodular goitre and solitary adenomas. Ultrasound or isotope scans are generally unhelpful in the investigation of Graves' disease, except in postpartum thyrotoxicosis, when it may be necessary to differentiate Graves' disease from self-limiting postpartum thyroiditis. In a toxic multinodular goitre, isotope scanning will usually show multiple hot areas, and will help determine the presence and extent of retrosternal extension. In a solitary toxic adenoma, an isotope scan will demonstrate the hot nodule with suppression of the extranodular thyroid tissue. Differentiation between a solitary toxic adenoma and a toxic multinodular goitre can help plan the extent of thyroidectomy. Fine needle aspiration (FNA) is usually used to exclude an associated thyroid neoplasm in the presence of a solitary or dominant thyroid nodule.

Treatment of Graves' disease

Antithyroid drugs

These are the mainstay of initial treatment. Although the hyperthyroid state can usually be well controlled, relapse rates are around 50–60%. Male sex, large goitres, high levels of thyroid hormones, high levels of TRAbs and smoking increase the risk of relapse. The thionamides (carbimazole, propylthiouracil and methimazole) block the organification (iodination) of the tyrosine residues on the thyroglobulin molecule, by inhibiting the enzyme thyroid peroxidase. Carbimazole is converted to its active constituent methimazole (which is primarily used in the USA). The traditional dosing regimen is 20–40 mg of carbimazole (100–150 mg of propylthiouracil) to render patients euthyroid and maintaining low-dose treatment for 12–18 months. An alternative treatment regimen ('block and replace' regime) involves using a much higher dose of antithyroid drugs along with thyroxine to maintain euthyroidism. The advantage of this latter regimen is faster control of hyperthyroidism with less frequent use of TSH measurements; the disadvantage being a higher incidence of side effects. Side effects of these drugs include rash, arthralgia and gastrointestinal disturbances; rarer but more serious side effects include agranulocytosis, vasculitis and hepatitis. Treatment with recombinant granulocyte colony-stimulating factor has been reported to shorten recovery time in patients with thionamide-induced agranulocytosis.

Radioactive iodine

RAI uses the high-energy particles of ^{131}I to ablate thyroxine-producing follicular cells. It is the treatment of choice in elderly patients, patients with small glands or those who are unfit for surgery. Contraindications to RAI include pregnancy (RAI freely crosses the placenta and activity in the maternal bladder causes fetal irradiation), breastfeeding (both iodine and pertechnetate are excreted in breast milk) and patients with severe toxicity (as patients may develop thyroid storm if markedly toxic and should be pretreated with beta-blockers). Carbimazole is stopped 48 hours before and restarted 3–5 days after RAI. The usual dose of ^{131}I given is 500–750 MBq and the maximum effects of treatment occur 3–4 months after the dose of RAI. Hypothyroidism is an almost inevitable consequence of treatment (see Hypothyroidism). RAI may exacerbate clinically evident ophthalmopathy and systemic corticosteroids may be indicated to reduce this risk. Although debatable, some studies suggest a small increase in cardiovascular and cancer-related mortality.

Surgery

Traditionally, the standard surgical treatment in Graves' disease was a subtotal thyroidectomy, the aim of which was to leave behind a small volume of thyroid tissue in each lobe so as to leave the patient euthyroid. However, the drawbacks were a risk of recurrent thyrotoxicosis (around 10%) and the observation that up to 70% of patients will develop hypothyroidism at 10 years. The current trend therefore is to perform a near total or total thyroidectomy and immediate institution of thyroxine postoperatively. This serves to eliminate the disease with negligible recurrence rates. There are concerns that the complications such as RLN damage and hypoparathyroidism may be higher with the more radical operation, but this has not been shown to be the case in experienced hands.

Treatment of toxic adenoma/multinodular goitre

For patients with toxic adenoma and toxic multinodular goitre, surgery and RAI are effective and safe treatment options. Factors favouring surgery include contraindications to RAI, large or compressive goitres, indeterminate or suspicious nodules, nodules with retrosternal extension and coexisting hyperparathyroidism needing surgery. Factors favouring the use of RAI include elderly, unfit patients, patients with small nodules and previous neck surgery. Occasionally, elderly patients unfit or unwilling to undergo either surgery or RAI may be maintained on long-term low-dose antithyroid drugs.

Thyroid nodules

Epidemiology

Nodular thyroid disease is common and the incidence increases with age. The prevalence of palpable nodules in areas of adequate iodine intake is 4–7% but autopsy studies and ultrasonography have shown the true figure to be much greater with around 50% of adults having nodules. In areas of severe iodine deficiency, the prevalence can be as high as 90%.

A wide range of benign and malignant pathology can account for patients presenting with a thyroid nodule(s). Malignant diseases are covered later, but the list of benign pathology includes:

- colloid or multinodular goitre (may present as solitary/dominant nodule)
- thyroiditis
- simple or haemorrhagic cysts
- follicular adenomas
- thyroid abscess.

In a multinodular goitre (Figure 18.3), enlargement of the thyroid tends to start as diffuse hyperplasia of the gland, with subsequent areas of focal hyperplasia (which may be dependent on TSH stimulation) and areas of regression and colloid degeneration; a multinodular goitre may be non-toxic (euthyroid) or toxic (Plummer disease). The aetiology of a multinodular goitre is poorly understood, but is thought to be multifactorial: genetic predisposition to dyshormonogenesis, iodine deficiency, drugs (notably amiodarone and lithium), ingestion of brassica vegetables and autoimmunity. Thyroid cancer, however, is rare, affecting around four in 100 000 individuals per year and constituting 1% of all malignancies.

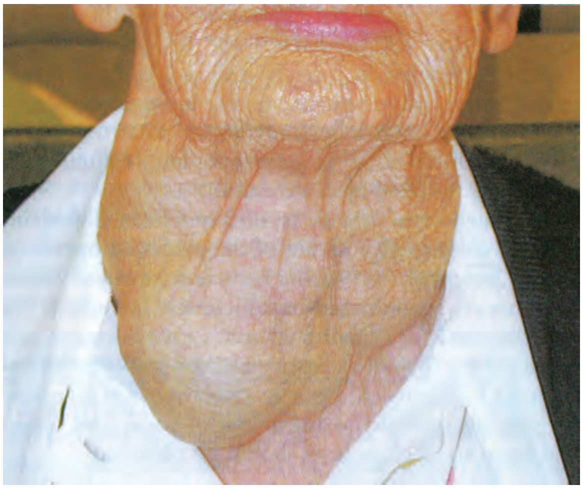


Figure 18.3 Patient with longstanding multinodular goitre.

The difficulty for the clinician and patients in the management of nodular goitre is to strike the right balance that avoids overtreatment of common benign nodular disease (Figures 18.4 and 18.5) and enables early detection and treatment of thyroid cancer. This is compounded by the need for diagnostic excision in the all too common scenario of indeterminate or inconclusive cytology.

Clinical features

The majority of thyroid nodules are asymptomatic and are found fortuitously by the patient or on routine clinical examination. The lumps are largely painless, but the sudden presentation of a painful swelling is almost pathognomonic of haemorrhage into a simple colloid nodule. Patients with thyroiditis sometimes have local discomfort and tenderness. Compressive symptoms such as difficulty in breathing and/or swallowing are usually seen in very large, multinodular goitres. These goitres can cause significant deviation and/or tracheal compression; the latter may be associated with stridor. Compressive or obstructive

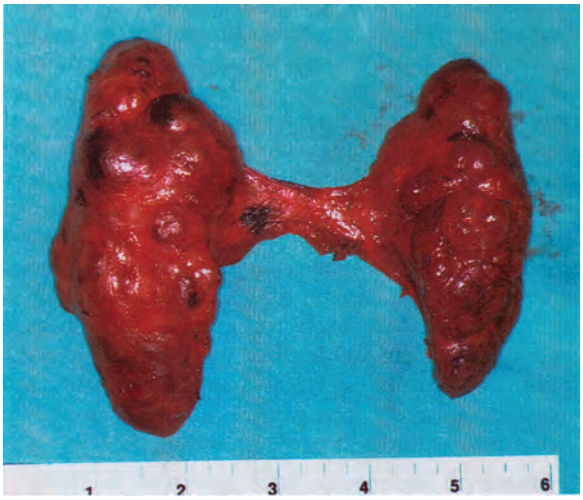


Figure 18.4 Dominant nodule in the right upper thyroid lobe of a multinodular goitre.

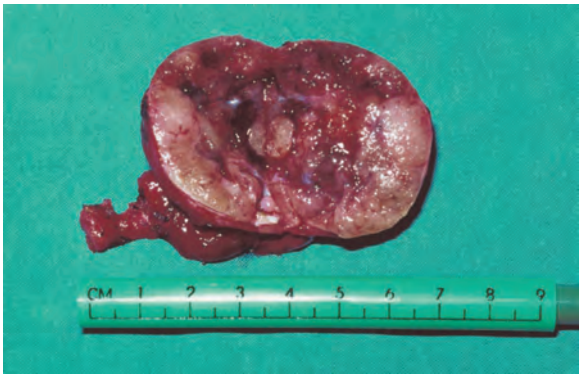


Figure 18.5 Solitary nodule – benign.

symptoms are more common when the goitre grows posteriorly or in a retrosternal direction; around 10% of multinodular goitres will have a retrosternal extension. These goitres may cause compression of large veins at the thoracic inlet leading to dilated veins in the head, neck and upper limbs. Elevation of the arms in conjunction with deep inspiration can often precipitate engorgement of veins in the upper half of the body, flushing and respiratory distress in these patients (Pemberton's sign). Rapidly worsening symptoms especially in association with voice change can be a sign of malignancy, but a malignant tumour can also be extremely slow growing and be present for many years before being discovered. Although the vast majority are benign, features of hyperthyroidism and hypothyroidism should be elicited in all these patients (see previous sections). Nodules in patients with a past history of neck irradiation and a family history of thyroid malignancy clearly have a high likelihood of being malignant. The very young and elderly are at increased risk for malignancy. There is an increased incidence of follicular cancer in iodine-deficient endemic goitrous areas and an increase in papillary cancer in iodine-rich regions. Solitary nodules convey a greater risk of malignancy in males, and a family history of endocrine disease suggests the possibility of medullary thyroid carcinoma. Papillary carcinoma may also be familial and has been described with familial adenosis polyposis (Gardner syndrome) and also ataxia–telangiectasia. Consistency can be misleading as, although a hard fixed nodule is likely to be malignant, a benign colloid nodule can also be hard with dystrophic calcification. Associated cervical lymphadenopathy and features of RLN palsy are highly suggestive of malignancy (Box 18.1). However, associated lymph nodes could be reactive (especially in presence of thyroiditis) and nerve palsy could occasionally be secondary to compression by a longstanding, large multinodular goitre.

Investigations

Fine needle aspiration cytology

Fine needle aspiration cytology (FNAC) is now a routine and important investigation in patients presenting with a solitary thyroid nodule or a dominant nodule in a multinodular goitre. It may be a curative procedure for simple thyroid cysts, although patients with recurrent cysts and those with residual solid areas should be reaspirated or considered for surgery. FNA is, however, not required in patients with multinodular goitre without a dominant nodule and also in patients with a diffuse goitre who are hyperthyroid (as in Graves' disease). Cytological assessment may be performed using a wet-fixed or air-dried preparation or alternatively utilizing a cell block method in which the thyroid architecture is preserved. The authors perform several passes of the needle for each nodule and find good patient compliance with few complications. The procedure can also be easily repeated (Figures 18.6–18.9).

In the UK, the cytology report is usually classified into one of the following categories:

- Thy1 (inadequate or non-diagnostic)
- Thy2 (benign or non-neoplastic)
- Thy3 (indeterminate, possibly neoplastic)
- Thy4 (suspicious of malignancy)
- Thy5 (diagnostic of malignancy).

Thy1 and Thy2 categories are sometimes suffixed as Thy1c and Thy2c, respectively, to indicate the possibility of a cystic lesion. Recently, the Thy3 category has been subdivided into Thy3a to indicate a lesion with atypical features of uncertain significance and Thy3f to indicate the likelihood of a follicular neoplasm. Often, a confident diagnosis of colloid nodule (Thy2) or thyroiditis (Thy2), papillary, medullary and anaplastic carcinoma (Thy5), lymphoma (Thy5) and even

BOX 18.1 Risk factors for thyroid malignancy

- Age – a new thyroid nodule in patients <20 years or >50 years
- Male sex
- Clinical features – consistency, fixation, size, solitary versus multiple nodules
- A history of head and neck irradiation
- Familial history of thyroid malignancy of multiple endocrine neoplasia (MEN) 2
- Recurrent laryngeal nerve palsy
- Cervical lymphadenopathy

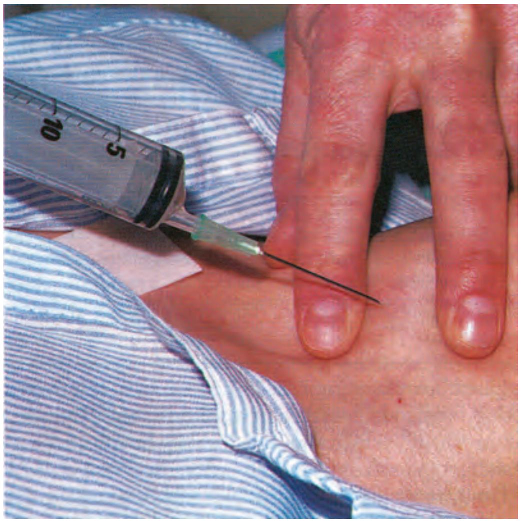


Figure 18.6 Fine needle aspiration: fixation of a nodule with the index and middle fingers prior to aspiration.

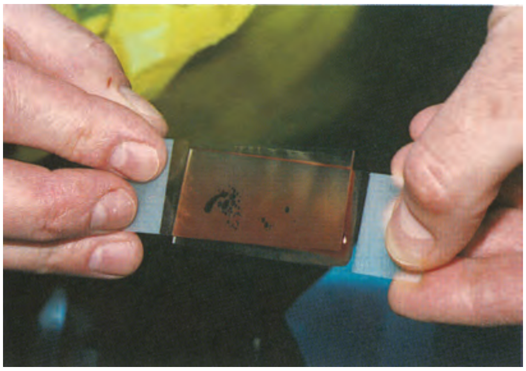


Figure 18.7 Fine needle aspiration: the aspirate is spread evenly between two slides.

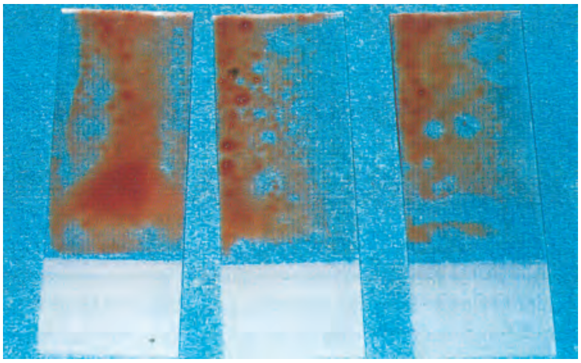


Figure 18.8 Fine needle aspiration: the specimen is to be air-dried before sending to the laboratory.

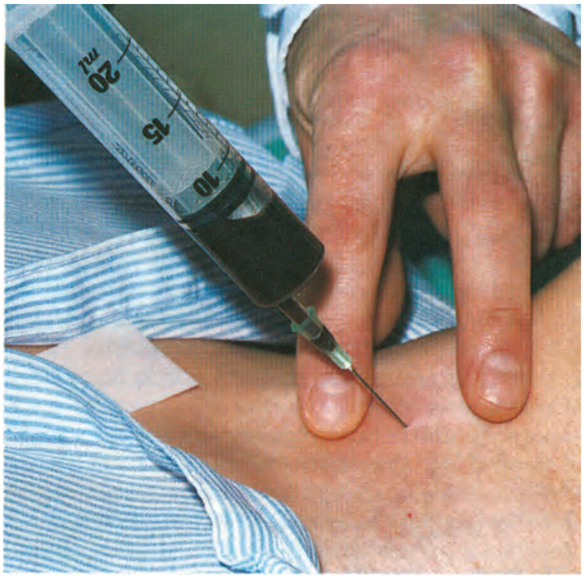


Figure 18.9 Fine needle aspiration cytology: a diagnostic and sometimes also a therapeutic procedure (as seen here with aspiration of a thyroid cyst).

metastatic deposits (Thy5) can be made. A major limitation of the technique is the evaluation of follicular lesions where histology is required to differentiate benign follicular lesion from carcinoma; the latter diagnosis is dependent upon the presence of capsular and vascular invasion. Core biopsy for these follicular lesions has been suggested by some authors, but increases the risk of haematoma and may still be inadequate for a definite histological diagnosis.

The positive and negative predictive values of a clear FNA result are over 98%. However, inadequate specimens should lead to repeat aspiration. Surgery may sometimes be indicated in a patient with repeated inadequate cytology. Benign lesions may be managed conservatively and FNA repeated to confirm the diagnosis. Clinical suspicion, increasing size of nodules, compressive symptoms and associated hyperthyroidism may be indications for surgery in nodules with benign cytology. Patients with indeterminate, suspicious and malignant results should undergo a thyroidectomy. The diagnosis of indeterminate and suspicious lumps may in the future be improved with immunocytochemical techniques.

Blood tests

These should be requested to measure TSH and free T_4 levels. The majority of patients with a thyroid nodule are euthyroid but coexisting thyroid dysfunction may point to the underlying pathology. For example, a hyperthyroid patient with a solitary nodule suggests a benign toxic (autonomous) nodule, whereas hypothyroidism may indicate nodular Hashimoto disease possibly with lymphomatous change. The presence of thyroid dysfunction will also influence further management. Patients with hyperthyroidism who need surgery would need to be made euthyroid with antithyroid drugs. In patients with a positive family history of MTC, serum calcitonin should be measured to aid diagnosis. This is also used for monitoring following treatment of MTC.

Imaging

High-resolution ultrasonography is now often performed as part of the diagnostic work-up of a thyroid nodule. It is sensitive in identifying impalpable nodules as small as 0.3 mm in diameter. Ultrasound detects multinodularity in around 50% of all nodules thought to be solitary on examination. It differentiates cystic from solid lesions and can identify associated lymphadenopathy, which could be a marker for malignancy. Other ultrasound features of malignancy include nodules with spiculated margins, hypoechogenicity and microcalcifications, but none of the findings are accurate enough to dictate clinical management. Its routine use in all patients presenting with a thyroid nodule or a multinodular goitre is probably unnecessary. An ultrasound is, however, quite useful in specific clinical scenarios. Ultrasound is used to characterize and guide FNA of a barely palpable nodule or an impalpable lump detected incidentally on other imaging [such as CT or positron emission tomography (PET)]. For patients being managed conservatively, ultrasound may also help in monitoring nodule size during follow-up. In the follow-up of patients with thyroid cancer, neck ultrasound helps to identify locoregional recurrence.

Chest radiograph, CT and MRI have little role in the differentiation of benign and malignant lumps, but do help to determine the presence and degree of retrosternal extension and extent of tracheal deviation or compression (Figure 18.10).

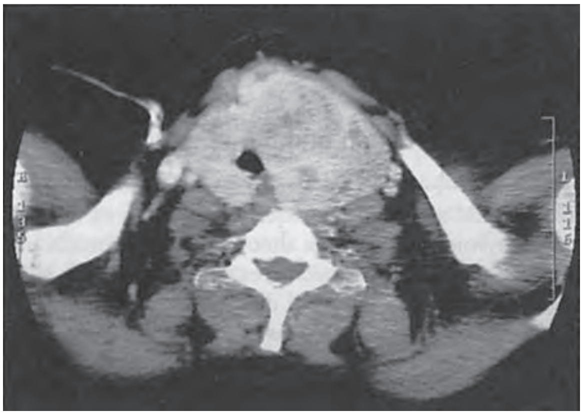


Figure 18.10 CT scan of the neck demonstrating tracheal deviation and compression.

Isotope scanning is not useful in differentiating between malignant and benign lesions and its role is now limited to patients with coexisting hyperthyroidism to aid in the diagnosis of a solitary toxic adenoma and to determine areas of hyperfunction in a patient with toxic multinodular goitre.

Treatment

This section will focus on the management of benign nodular disease. The treatment of the various forms of thyroid cancer will be covered in subsequent sections. An algorithm for the management of nodular thyroid disease is presented in Figure 18.11.

FNAC may be both a diagnostic and therapeutic tool for the management of simple thyroid cysts. However, surgical excision may still be required in a small number of patients in whom there is recurrent cyst formation or a suspicious residual nodule.

A long-term follow-up study of putatively benign thyroid nodules has demonstrated that just over a third of nodules disappear and most nodules reduce in size over a 10–30 year period (Kuma *et al*. 1992). However, 26% of enlarging nodules were found to be malignant. A further report by the same group (Kuma *et al*. 1994) on clinical re-examination, FNAC and ultrasound-guided FNAC to assess nodules over 9–11 years has clearly demonstrated that 99% of benign nodules remain benign, with the majority decreasing in size or disappearing during the follow-up period. The worrying clinical feature remains an increase in nodule size. Clearly there should be a high index of suspicion for lesions that increase in size during follow-up.

Surgery

The majority of multinodular goitres are benign and not associated with significant compressive symptoms and can therefore be treated conservatively. However, surgery is indicated for the following reasons:

- suspected or proven malignancy on FNA
- compression of the trachea or oesophagus
- significant recent growth of a dominant nodule (suggestive of malignancy)
- local neck discomfort
- cosmetic reasons.

The need for surgery is based on the overall assessment of symptoms, clinical risk factors and cytology. Indeterminate, suspicious or malignant cytology are clear indications for surgery. Recurrent cysts and solid nodules that repeatedly yield an inadequate sample on cytology are also indications for surgery. Other scenarios in which surgery is considered include patients with large nodules causing pressure symptoms (e.g. dyspnoea, dysphagia or choking sensation), nodules increasing in size on follow-up and increased patient anxiety especially in the context of a family history of thyroid cancer.

Surgery for nodular thyroid disease usually includes one of the following operations:

- hemithyroidectomy/lobectomy for unilateral disease
- isthmectomy for a solitary nodule confined to the isthmus
- total thyroidectomy for bilateral nodularity and solitary nodules with malignant cytology.

Nodule excisions and subtotal resections of either the lobe or the thyroid gland are no longer performed. In patients with a preoperative diagnosis of thyroid cancer, a prophylactic central node dissection is performed by many authors at the time of the total thyroidectomy. If suspicious/confirmed lateral lymph node metastases are present, a lateral node dissection is also done (see below).

During a hemithyroidectomy/lobectomy for unilateral disease with indeterminate/suspicious cytology, the surgeon palpates the contralateral lobe through the strap muscles. If the other lobe is grossly normal and there is no associated lymphadenopathy,

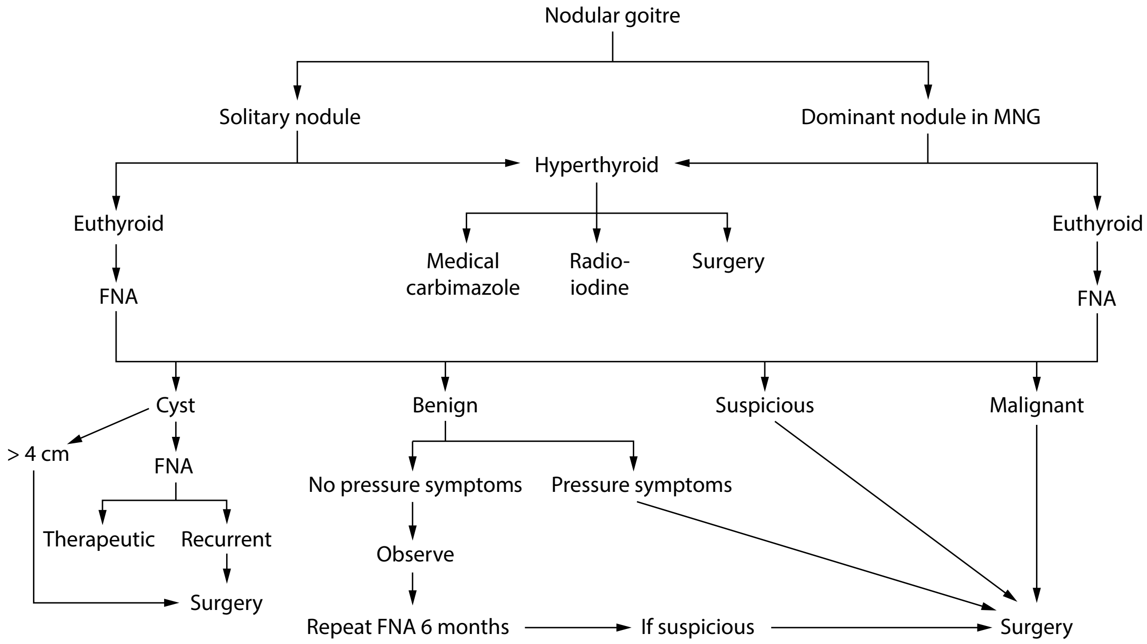


Figure 18.11 Algorithm for the management of a patient with a thyroid nodule. FNA, fine needle aspirate; MNG multinodular goitre.

a total thyroid lobectomy removing isthmus and pyramidal lobe (if present) is performed preserving the parathyroid glands, the EBSLN and the RLN. This is a safe procedure with low morbidity when performed in experienced hands. Frozen section may occasionally be useful in confirming malignancy leading to total thyroidectomy and avoiding a second operation. If there is any doubt (particularly with follicular lesions) then the neck is closed and formal paraffin histology is awaited.

Non-surgical options of treatment of benign non-toxic goitre

Several non-surgical options have been explored in an attempt to avoid surgery in patients with benign non-toxic goitre. RAI is a particularly useful alternative to surgery in patients who are unfit for surgery, unwilling to consider surgery and in those with recurrent disease. RAI may reduce goitre volume by around 50% and improve compressive symptoms. The drawbacks include transient hyperthyroidism, occasional increase in goitre size, late onset hypothyroidism and the theoretical increased risk of malignancy. The lack of a consistent effect may be related to low or patchy uptake in many patients – an argument for higher dose recommended by some authors. More recently, recombinant TSH has been used to increase uptake at lower doses and may prove more effective than RAI alone.

Thyroxine and iodine supplementation have been tried but are not routinely recommended in view of inadequate response rates, side effects (hyperthyroidism, cardiovascular and skeletal problems) and potential for regrowth following cessation of treatment. Percutaneous injection of ethanol has also been tried, but its use is largely experimental.

Thyroid cancer

Thyroid cancer is uncommon; the age-standardized incidence rate in the UK is around three per 100 000 per year. The incidence in the USA is twice that in Europe; the reasons for this are not clear but could be related to the increased use of imaging resulting in the detection of early stage, good prognosis cancers. Most thyroid cancers fall under the category of differentiated thyroid cancer (DTC), among which papillary thyroid cancer (PTC) makes up the vast majority. The various types of thyroid cancer are shown in Figure 18.12, and the risk factors for thyroid malignancy are shown in Box 18.1.

Differentiated thyroid cancer

Epidemiology

These cancers arise from follicular epithelial cells of the thyroid gland. The incidence of DTC in general and PTC in particular has more than doubled in the last 30 years. Interestingly, however, the incidence of follicular thyroid cancer (FTC) has not increased; this is at least partly due to a recent increase in the number of follicular cancers being classified as follicular variant PTC.

Papillary thyroid cancer

PTC accounts for around 80% of all malignant thyroid tumours. The histological diagnosis rests on characteristic nuclear features. Typical PTC shows a complex papillary architecture. It is also classically described as 'psammoma bodies', which are spherical areas of concentrically laminated calcifications in stroma or lymphatic spaces. Many variants of PTC exist and these include encapsulated, microcarcinoma (<1 cm), follicular, oncocytic, clear cell, diffuse sclerosing, tall cell, columnar cell, solid and cribriform variants. PTC typically spreads by the lymphatic route and occult lymph node metastases are thought to be present in around half of patients without any evidence of lymphadenopathy. The prognosis of these cancers in general is very good with >90% 10-year survival. Some subtypes such as the diffuse sclerosing, tall cell and columnar cell variants carry a worse prognosis.

Follicular thyroid cancer

This accounts for around 10% of all thyroid tumours. These tumours also arise from follicular cells, but lack the nuclear features of PTC. Iodine deficiency is thought to play an important role in its development. The cellular appearances are indistinguishable from follicular adenomas and therefore, as explained earlier, cytology is inadequate to differentiate between follicular adenomas and carcinomas. The types of FTC include minimally invasive follicular carcinomas, widely invasive follicular carcinomas, oncocytic follicular carcinomas and clear cell carcinomas. These cancers are often multifocal and spread predominantly by the haematogenous route to organs such as lung and bone. Lymph node metastases are uncommon. Prognosis is slightly worse than PTC; survival is around 85% at 10 years.

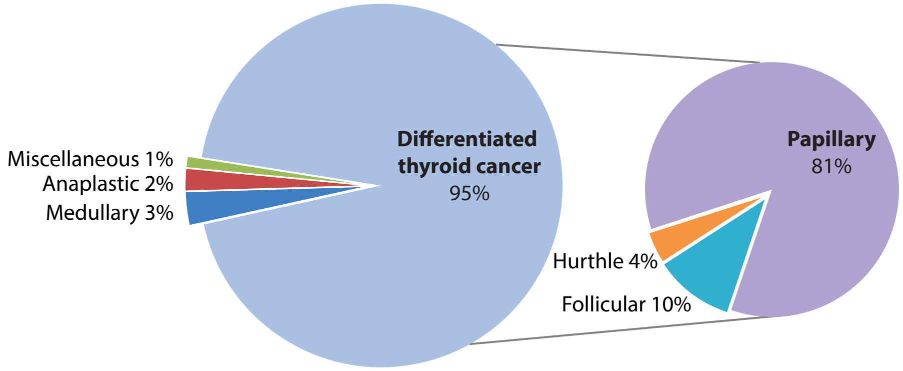


Figure 18.12 Thyroid cancer types and subtypes.

Hurthle cell cancer

Hurthle cell cancer (HCC) is a distinct variant of FTC, although Hurthle cells can be found in a variety of other neoplastic and non-neoplastic conditions. Hurthle cells typically have abundant altered mitochondria that fill up the cytoplasm resulting in the characteristic eosinophilic granular cytoplasm. In HCC, the lesion is composed predominantly (>75%) of Hurthle cells. These have a poorer prognosis than other FTCs and are slightly resistant to radioiodine. In variance with other FTCs, lymph node metastases occur in around 20% of patients with HCC.

Poorly differentiated cancers

These lie in the spectrum between differentiated and undifferentiated (anaplastic) cancers and arise from the follicular epithelium. These are often infiltrative and have necrotic areas and obvious vascular invasion. Insular, trabecular and solid patterns are recognized.

Mixed tumours

These are also recognized, in which varying proportions of the above tumour types are seen.

Prognosis

The prognosis of DTC is in general excellent with overall cure rates of around 90%. In general terms, 80–90% of patients will fall within the 'best' prognostic group, in whom disease-specific death occurs in no more than 2% of patients. In the worst prognosis groups, the 20-year mortality rate is around 75–95%. Some subtypes (tall cell, diffuse sclerosing and columnar varieties of PTC, HCC) and some factors (such as male sex, age >45 years at diagnosis, incomplete resection, extrathyroidal spread, presence of distant metastases) are associated with a worse prognosis and have higher recurrence rates. Several classification systems have been proposed to stratify risk and include the following:

- AMES: Age, Metastases, Extent of primary tumour, Size of presentation.
- AGES: Age, Grade, Extent of primary tumour, Size of tumour.
- MACIS: Metastases, Age, Completeness of resection, Invasion of extrathyroidal tissues, Size of tumour.
- TNM staging (5th edition):
  - *Tumour*
  - Tx: Primary tumour not identified or assessed
  - T0: No evidence of primary tumour
  - T1a: Tumour 1 cm, limited to thyroid
  - T1b: Tumour >1 cm and 2 cm
  - T2: Tumour >2 cm and 4 cm, limited to thyroid
  - T3: Tumour >4 cm, limited to the thyroid or with minimal extrathyroidal extension to strap muscles of perithyroidal soft tissue
  - T4a: Tumour invading subcutaneous soft tissues, larynx, oesophagus or RLN
  - T4b: Tumour invading prevertebral fascia, carotid sheath or mediastinal vessels
  - *Node*
  - Nx: Regional lymph nodes not assessed
  - N0: No regional lymph node metastasis
  - N1a: Level VI node metastases
  - N1b: Metastases to unilateral, bilateral or contralateral cervical nodes (levels I, II, III, IV or V), retropharyngeal nodes or superior mediastinal nodes (level VII).
  - *Metastasis*
  - Mx: Metastasis not assessed
  - M0: No distal metastases
  - M1: Distal metastases present.

Clinical features and investigations

These have largely been discussed in the section on thyroid nodules. The diagnosis of DTC may be made on clinical grounds, by thyroid cytology or following a thyroidectomy. Most cases are diagnosed following a lobectomy or total thyroidectomy for nodules that are apparently benign, indeterminate or suspicious on cytology. In patients with cytology diagnostic of cancer, an ultrasound (with or without CT/MRI in some centres) may help to detect abnormal regional lymph nodes. This will enable an appropriate level lymphadenectomy to be done at the time of thyroidectomy. In patients with clinically palpable lymphadenopathy or locally advanced disease (abnormal regional lymph nodes or a locally infiltrative tumour), CT or MRI of the neck is essential to ascertain the extent of extrathyroidal spread and lymphadenopathy.

Treatment

Thyroidectomy is the primary treatment for most patients with DTC. Most surgeons recommend total or near total thyroidectomy for all patients with DTC but unilateral lobectomy and isthmectomy may be adequate for some *low-risk* patients with DTC such as:

- patients with papillary microcarcinoma (<1 cm)
- minimally invasive follicular cancer without vascular invasion in young patients
- the encapsulated follicular variant of PTC.

In the significant proportion of patients in whom the diagnosis is only made following an isthmectomy or hemithyroidectomy and a completion thyroidectomy is required, this should be performed within 7 days or after 3 months have elapsed from primary surgery to minimize risks of morbidity associated with postoperative scarring.

Lymph node dissection

The need for and extent of lymphadenectomy depends on the type of cancer (PTC and HCC are associated with lymph node spread) and the presence of palpable lymphadenopathy. Lymph nodal spread is usually first to the central compartment (levels VI and VII) and then to levels III, IV and V in the lateral compartment. However, occasionally skip metastases occur. Patients with obvious lymph node metastases need a central compartment node dissection and lateral compartment node dissection on the involved side. Levels I and II are usually spared but if involved should be dissected as well. The role of prophylactic central compartment dissection in patients with PTC and HCC without obvious nodal disease is more controversial. Proponents of routine prophylactic central neck dissection point to the high incidence of occult lymph node metastases in PTC and demonstrate low rates of side effects such as RLN damage and hypoparathyroidism in experienced hands. Opponents of prophylactic neck dissection argue that occult nodal metastases are unlikely to affect long-term outcomes and can be dealt with if and when identified on follow-up. The authors adopt a selective policy and perform a prophylactic central node dissection in patients with high-risk features such as male sex, age >45 years, tumour >4 cm in diameter, extracapsular spread and extrathyroidal invasion.

Radioactive iodine

An ablative dose of RAI (^{131}I) is used as standard adjuvant treatment following total or near total thyroidectomy for DTC. The treatment is based on the premise that thyroid cancer cells in DTC are iodine avid and would be ablated by RAI. RAI is not recommended for patients undergoing only a hemithyroidectomy for low-risk disease. The ablation of any residual thyroid tissue also facilitates the use of thyroglobulin as a tumour marker to screen for residual and recurrent disease. Thyroid hormones are usually withdrawn in preparation for RAI treatment to allow TSH levels to rise and increase iodine uptake by thyroid tissue. Patients receiving RAI within 3–4 weeks after thyroidectomy do not need treatment with thyroid hormones. However, for patients in whom RAI treatment may be delayed or for patients needing further RAI treatment for residual or recurrent disease, treatment with T_3 is started and this is withdrawn for 2 weeks before RAI treatment. Patients are also advised a low-iodine diet for 2 weeks prior to treatment. Although preablation diagnostic ^{123}I is now not routinely used, it may help in assessment of residual thyroid surgery if completeness of resection is not known. Postablation diagnostic scans are often carried out to assess the effectiveness of ablation.

Thyroxine suppression

Normal thyroid cell differentiation and proliferation is TSH dependent. TSH receptors are also present in DTC cells. Thyroxine suppression treatment is based on the principle that high doses of thyroxine would suppress TSH levels and thereby proliferation of thyroid cancer cells. Patients receive lifelong thyroxine, at doses necessary to suppress TSH to undetectable levels. T_3 may be used if follow-up radioactive iodine scans or ablation are required, as it has a shorter half-life than thyroxine, and only requires to be stopped 2 weeks prior to RAI use. Long-term effects of suppressive doses of thyroxine include cardiovascular morbidity and osteoporosis.

External beam radiotherapy

External beam radiotherapy to the neck is used only in patients with extensive extrathyroidal disease and in patients whose tumours are not radioiodine avid.

Follow-up of patients with differentiated thyroid cancer

Recurrent thyroid cancer may occur soon after initial therapy or years later. Patients with thyroid cancer should be followed up lifelong in a multidisciplinary thyroid cancer clinic. A history and examination looking for evidence of recurrent neck lumps and features of thyrotoxicosis and measurements of TSH and thyroglobulin (including thyroglobulin antibody) levels should be done at every visit. Calcium levels should also be measured in patients who have persistent hypothyroidism following surgery. A follow-up diagnostic radioactive iodine scan may be needed in patients with high-risk disease or if the serum thyroglobulin levels are high/rising, which should be minimal or undetectable in patients who have undergone definitive treatment (surgical and radioactive iodine). Diagnostic and ablation scans both require hormone withdrawal, although more recently recombinant TSH has been used as an alternative. Ultrasound scanning of the neck may be necessary in patients with palpable lumps and rising thyroglobulin levels, although it also used routinely for follow-up in some centres. 18-Fludeoxyglucose–PET scanning can help detect recurrent disease in scan-negative patients with rising thyroglobulin levels.

Recurrent differentiated thyroid cancer

Recurrence may be local or regional (in the neck) or systemic. Treatment will depend on the subtype of DTC, the nature and extent of initial treatment and the site of recurrence. Surgery is the mainstay of treatment of local/regional recurrence. Surgery may be followed by RAI scanning with/without ablation and, in some cases, external beam radiotherapy. Distant metastases will usually occur in the lungs or skeletal system. Although distant metastases cause cancer death in 10–15% of patients with DTC, they are compatible with long-term survival. They may present as a result of local symptoms, neurological complications or rising thyroglobulin levels. Radioactive iodine is the mainstay of treatment for recurrent disease. Palliative surgical procedures may be appropriate if there are orthopaedic/spinal complications. It should be remembered that metastases may be solitary. Remission will occur in 50% of patients with distant metastases that take up radioiodine (10 years survival 25–40%).

Medullary thyroid cancer

Epidemiology

MTC accounts for up to 5% of all thyroid cancer. This tumour arises from the parafollicular C cells of the thyroid, which have a neural crest origin. Around a quarter of these cases are familial. The disease may occur in one of four clinical settings:

1. sporadic
2. familial – MEN 2A
3. familial – MEN 2B
4. familial non-MEN-associated MTC.

The mortality rate for MTC exceeds that of DTC; overall, the 10 year survival rate is around 75%. Factors that indicate a poor prognosis include age >40 years at presentation, male sex, extrathyroidal spread, nodal involvement, metastases, tumour aneuploidy, negative amyloid staining and familial disease. However, long-term survival of patients with metastatic disease is common in MTC.

Clinical features and investigation

Sporadic MTC usually presents as a thyroid nodule and/or lymph node enlargement. Associated symptoms include those secondary to airway/oesophageal compression, pain, diarrhoea and rarely Cushing syndrome owing to gut peptide or adrenocorticotrophic hormone release by the tumour. The diagnosis may be confirmed on thyroid cytology or following lobectomy for a thyroid nodule. In all cases, family history for thyroid cancer/phaeochromocytoma should be determined. The absence of a family history does *not* preclude an apparently sporadic MTC being the index case of genetically determined disease.

If the diagnosis of MTC is made on cytology, preoperative investigations should include basal calcitonin and carcinoembryonic antigen (CEA), ultrasound of the neck to identify multiple thyroid lesions (a marker of familial disease) and lymph node enlargement. CT/MRI may identify mediastinal node involvement. Genetic screening (on a venous blood sample) for a *Ret* proto-oncogene mutation is required in all patients to exclude familial disease; but this can wait until after surgery.

Phaeochromocytoma must be excluded prior to operation in all cases by a normal 24 hour urine collection for catecholamines/metanephrines.

In families affected by or likely to be affected by genetically determined MTC, screening for *Ret* mutations in individuals at risk should be performed. Prophylactic thyroidectomy is indicated in family members without clinically apparent disease but who are carriers of the germline *Ret* mutation. The different *Ret* mutations are associated with different degrees of susceptibility and disease aggressiveness and recent recommendations on the age at which prophylactic thyroidectomy should be performed in children positive for this mutation are based on the site of the mutation.

Treatment

The current standard operation in the absence of lymph node metastases is total thyroidectomy and central compartment neck node dissection. Biopsy and frozen section examination of any enlarged jugulocarotid lymph nodes from either side of the neck should be performed. If these nodes are positive they should be removed and a lateral neck dissection should be done. If lymph node metastases are detected preoperatively in the lateral compartment, a selective node dissection of this compartment should be done at the same sitting. If there is evidence of involvement of anterior/superior mediastinal node involvement at presentation, these nodes should be cleared (sternotomy would be required). In all cases the parathyroid gland should be identified and preserved. In MEN 2 patients only enlarged parathyroid glands should be excised. After surgery *replacement* doses of thyroxine are given. There is no indication for TSH suppression or radioactive iodine in the treatment of MTC.

Follow-up of patients with medullary thyroid cancer

Long-term follow-up of these patients is required. At review, basal calcitonin, CEA and TSH levels should be measured in addition to a history and physical examination. Detectable or raised calcitonin and/or CEA levels indicate residual or recurrent disease and warrant a search for locoregional and metastatic disease. This requires the use of high-resolution CT of the neck/chest/liver, and/or one or more of the isotope scans with pentavalant dimercaptosuccinic acid, ^{123}I-metaiodobenzyl guanidine or radiolabelled octreotide. Further treatment decisions (including surgical resection or observation) are made in a multidisciplinary forum and are influenced by the presence of symptoms, levels of tumour markers, findings on imaging, age and associated morbidity.

Patients with MEN 2 require biochemical testing to exclude a phaeochromocytoma and primary hyperparathyroidism at least on an annual basis.

Treatment of recurrent/metastatic disease

Surgery is the treatment of choice for local/regional recurrence. MTC is generally considered resistant to chemotherapy. The response to radiotherapy is generally poor but may be useful in patients with inoperable disease and symptomatic bone metastases. Diarrhoea may be severe and intractable in recurrent disease and should be controlled by the use of antidiarrhoeal agents including codeine phosphate and loperamide.

Undifferentiated thyroid carcinoma

Epidemiology

Undifferentiated or anaplastic thyroid carcinoma, in contrast to well-differentiated thyroid carcinoma, is a highly aggressive tumour, with around 70% of patients having metastases at the time of presentation. This is more common in areas with iodine deficiency and endemic goitre. Undifferentiated thyroid carcinoma may be classified as small cell, large cell or spindle cell, which may resemble sarcomas. Small cell carcinomas must be distinguished pathologically from lymphoma, which has a far more favourable prognosis.

Clinical features and Investigations

Diagnosis can usually be confirmed by a core biopsy under local anaesthetic, as FNA for cytology may not be diagnostic. Patients generally present with a large, hard, ill-defined cervical mass fixed to adjacent structures. There is often a history of a longstanding goitre. The sex incidence is similar and most patients present between 60 and 70 years of age. Occasionally, undifferentiated thyroid carcinoma may develop as a transformation of a previously treated well-differentiated thyroid carcinoma, which may have been in remission for a considerable time.

Treatment

Treatment of these patients is with a multimodal approach using surgery, radiotherapy and chemotherapy. Surgery is feasible and often necessary to relieve symptoms. Even in the absence of extrathyroidal spread, it is only considered a palliative procedure. Often only debulking is possible and a tracheostomy is required. Tracheal/oesophageal stenting should sometimes be considered if surgery is not successful or possible. External beam radiotherapy is usually given to these patients in an attempt to slow tumour progression. Response to chemotherapy is generally poor but regimes containing doxorubicin can give partial remission in around 20% of patients. Undifferentiated thyroid carcinoma is not responsive to ^{131}I therapy.

In some centres, preoperative chemoradiation followed by total thyroidectomy in patients has shown some promise. However, no standard treatment protocol exists currently and efforts are under way to investigate newer chemotherapeutic agents including tyrosine kinase inhibitors in an attempt to improve response. The prognosis remains very poor, with the majority of patients dead within a year of diagnosis.

Thyroid lymphoma

Epidemiology

Thyroid lymphomas are rare, occur more often in women (female to male ratio 3:1) and the incidence increases with age (most patients are aged >60 years). Thyroid lymphoma is often associated with a history of autoimmune thyroid disease (80% of patients).

Clinical features and investigations

The diagnosis is often suspected clinically by the history of long-standing goitre/hypothyroidism and a rapidly enlarging neck mass with a minority complaining of compressive symptoms. FNA is often inconclusive but usually shows plenty of lymphocytes. A core biopsy is needed to confirm the diagnosis and allow immunohistochemical subtyping of the lymphoma. Most thyroid lymphomas are mucosa-associated lymphoid tissue lymphomas and diffuse large -cell lymphomas. CT scanning often shows homogeneous thyroid enlargement without invasion of adjacent structures. CT of the chest, abdomen and pelvis is also done as part of staging of the lymphoma. Other investigations for lymphoma include full blood count, $_2$-microglobulin and bone marrow biopsy.

Treatment

Some patients will present with acute airway obstruction. Intravenous steroids can achieve rapid resolution of symptoms but, wherever possible, it should be given after a tissue diagnosis has been obtained. External beam radiotherapy may also be of value in relieving acute symptoms.

There is no evidence that surgery offers any benefit to patients with lymphoma. Following diagnosis and treatment of upper airway symptoms the patient should be referred to an oncologist. Staging with CT scanning, liver function tests and full blood count is performed. Treatment for this disease includes chemotherapy (the 'CHOP' (consists of cyclophosphamide, doxorubicin, vincristine and prednisolone) regimen) and/or radiotherapy.

Thyroidectomy

It was not until the early twentieth century that thyroidectomy became a safe and acceptable operation with the advent of general anaesthesia, antisepsis and haemostatic techniques. Theodore Kocher of Berne, Switzerland, was the chief protagonist of these methods and, for his lifetime devotion to the development of safe thyroid surgery, was awarded the Nobel Prize in 1909, by which time the previous high mortality had fallen to <1%. Further advances by William Halsted, Charles Mayo, George Crile and Frank Lahey led to techniques that remain the basis of safe thyroid surgery and continue to be practised today by trained endocrine surgeons.

Preoperative preparation

As with all surgical procedures, a full and written informed consent should be obtained after explaining the need for the operation, the implications of having the procedure, the risks of complications (see below), alternative treatment options and any other relevant details the patient might wish to discuss. A higher risk of complications should be emphasized when re-exploration or cancer surgery is performed.

Preoperative preparation should include laryngoscopy to exclude pre-existing unilateral nerve palsy, especially if the patient has undergone previous thyroid surgery. A serum calcium level is also routinely obtained as baseline to compare postoperative calcium levels with.

Surgical principles and approaches

Removal of one or both thyroid lobes, isthmus and additional thyroid tissue, such as the pyramidal lobe, as well as identification and preservation of the laryngeal nerves and parathyroid glands are the key objectives of thyroid surgery. A detailed knowledge of surgical anatomy, meticulous dissection and care to avoid bleeding are crucial to the achievement of these objectives. A number of different instruments and aids are now in use for dissection including the traditional 'tie and cut', bipolar diathermy, clips, ultrasonic dissection and thermal coagulation. However, these cannot replace careful dissection by experienced hands.

The standard approach to a thyroidectomy is through a collar incision in the neck. Endoscopic approaches have been described for small thyroid nodules (<3 cm in diameter), the primary aim of which is to avoid a neck scar. These procedures, however, involve the introduction of a camera and endoscopic instruments through incisions in the chest and axillae; have no proven superiority in terms of clinical outcomes; have a prolonged learning curve; and are not practised by the authors. Further description will only relate to the open approach. It is acknowledged that there will be significant variations in operating technique as practised by surgeons across the world; the following technique is the one that is used by the authors.

Operative steps

General anaesthesia with endotracheal intubation is deployed and the patient is placed supine on an operating table 15° head-up and the neck in near full extension with a sandbag (or inflatable bag) in the interscapular position.

Access

A collar incision is used two finger-breadths above the sternal notch extending to both sternomastoid muscles. The incision is extended through subcutaneous fat and platysma down to the deep fascia and, by a process of blunt and sharp dissection, this plane (anterior to the anterior jugular vessels) is extended superiorly to the level of the thyroid notch and inferiorly to the sternal notch, with the skin flaps then held apart using a self-retaining retractor. The strap muscles on either side are separated by dividing the deep fascia in the midline and retracted laterally. This should be as long as possible to enable full access to the operative field. Transverse division of the strap muscles is not routinely required but may be occasionally used for safe access to a large or vascular goitre. The deeper sternothyroid muscle is usual slightly adherent to the capsule and is separated from it with careful blunt and sharp dissection.

The middle thyroid veins, when present, are divided to increase access to the thyroid lobe, which is then delivered using traction from the index finger on a small swab over the lobe with the strap muscles retracted laterally. The lateral aspect of the thyroid lobe is now exposed from the superior pedicle superiorly to the lower pole inferiorly. The plane between the thyroid, larynx and oesophagus medially and the carotid artery laterally is opened by dividing the loose areolar tissue in this region; this increases the mobility of the thyroid lobe.

Superior pedicle

The upper pole is then retracted inferolaterally to open the space between the superior thyroid vessels and the cricothyroid muscle (Figure 18.13). The branches of the superior thyroid pedicle are divided at their entry into the thyroid on the surface of the capsule well away from the EBSLN, which usually runs along the surface of (or within) the cricothyroid but may also pass between the branches of the vessels where it is in danger of injury if mass ligation of the superior thyroid pedicle is carried out. The superior pole of the lobe may now be delivered partially into the midline. Further ligation of the small vessels entering the thyroid capsule may be required at this stage and care should be taken to avoid injury to the superior parathyroid gland.

The recurrent laryngeal nerve

Medial retraction on the lobe will now bring into view the deeper aspect of the middle third of the thyroid lobe and the adjacent junction between the inferior thyroid artery and the RLN (Figure 18.14). The RLN can be identified in three locations – upper (at the point of entry into the larynx at around the cricothyroid junction), middle (at the junction with the inferior thyroid artery) and lower (between the medial wall of the carotid artery and the lateral wall of the thymus). The nerve is seen following dissection of the overlying fascial layers and recognized as a white cord with an overlying vasa nervosum. The nerve lies in the tracheo-oesophageal groove (especially on the left) and runs upwards, passing anterior, posterior or through the branches of the inferior thyroid artery. However, its course is variable, especially on the right side. The right RLN lies in a more oblique direction. In around 1% of cases the right nerve may be non-recurrent and pass medially close to the inferior thyroid artery before turning to ascend to enter the larynx. These nerves also occasionally divide into branches before entering the larynx where the inferior cornu of the thyroid cartilage is a fairly constant landmark for its point of entry. The nerve is perhaps most in danger at its point of entry into the larynx as it passes through the suspensory ligament of Berry, where it often adopts a curving loop (Figure 18.14), and the nerve must be carefully identified in this region before dividing the suspensory fascia by staying close to the thyroid capsule at all times.

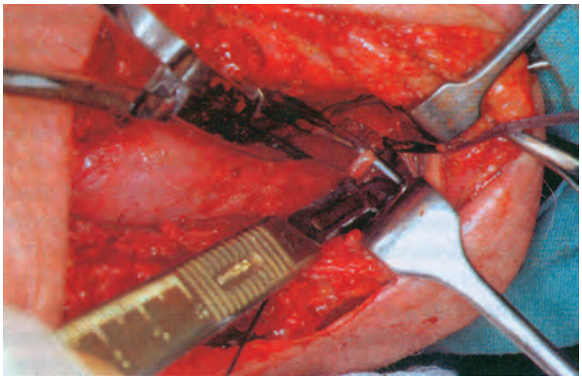


Figure 18.13 Division of the left superior thyroid artery.

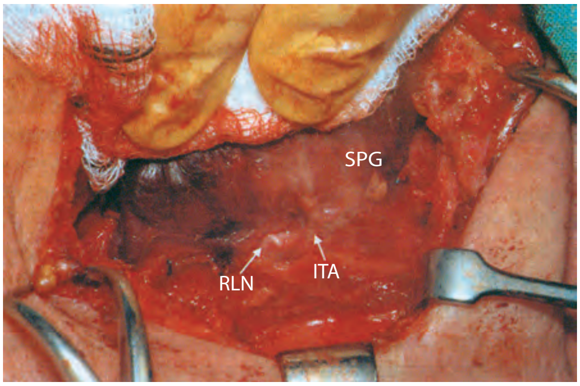


Figure 18.14 Exposure of the neurovascular intersection of the inferior thyroid artery (ITA) and recurrent laryngeal nerve (RLN). The superior parathyroid gland (SPG) can be found within a 2 cm radius cranial to the intersection, usually posterior to the RLN and thyroid gland. Note the looping of the RLN within Berry's ligament close to its insertion beneath the cricothyroid.

The parathyroid glands

The superior parathyroid gland is most commonly found superior to the inferior thyroid artery and posterior to the RLN, and the inferior gland often lies inferior to the inferior thyroid artery and anterior to the RLN. Attempts should be made to identify both the superior and inferior glands during lobectomy, but it may sometimes be appropriate to avoid an extensive dissection simply to identify the parathyroid glands as the dissection increases the risk of devascularizing the glands. In addition to the glands, care should be taken to preserve individual branches of the thyroid arteries supplying them. It is for this reason that the inferior thyroid artery is not ligated at its trunk, so as to preserve the small branches supplying the parathyroid glands. If the parathyroid glands are rendered ischaemic at thyroidectomy the individual gland may be excised and is minced into 1 mm cubes and autotransplanted into a pocket in the sternomastoid muscle.

Resection of the lobe

Dissection is continued by dividing the individual branches/tributaries of the main thyroid arteries and veins close to the thyroid capsule. Further dissection aims to divide the dense fascia binding the thyroid lobe to the trachea and larynx with particular attention to clipping and ligation near Berry's ligament, where troublesome bleeding may obscure the entry point of the RLN to the larynx. The mobilization is now complete and dissection is continued to include the isthmus and pyramidal lobe where present. The cut surface of the contralateral thyroid lobe is usually sutured with fine absorbable sutures to the tracheal fascia to obtain haemostasis.

Wound closure

The sandbag is now removed (or the inflatable bag can be deflated) from under the patient's spine and the neck space is examined for bleeding (occasionally a Valsalva manoeuvre may be performed by the anaesthetist). Haemostasis is secured and the wound is closed in layers (strap muscles, platysma and subcuticular layers) with 3/0 Vicryl or Monocryl. Drains are not required routinely, but may be useful in cases of significant oozing from the thyroid bed to prevent seromas after resection of very large goitres and if a neck dissection has also been done.

Total and subtotal thyroidectomy

In patients undergoing total thyroidectomy for cancer, bilateral multinodular disease or Graves' disease, the opposite lobe will be mobilized in a similar manner to that described above. Central compartment (with/without lateral compartment) node dissection may also be done at the same time (see below for details). Subtotal thyroidectomy, in general, is avoided, although it may be useful in Graves' disease if compliance with thyroxine replacement postoperatively is unlikely. In this procedure, a small remnant (usually 4–5 g of tissue) is left on each side of the trachea and sutured to the trachea with 3/0 Vicryl absorbable sutures to secure haemostasis. Some surgeons instead perform a unilateral total lobectomy leaving a single larger remnant on the contralateral side, which is an acceptable alternative strategy.

Retrosternal goitre

Ligation and division of the superior vessels is essential before any attempt is made to deliver the retrosternal component. This is achieved by introducing a finger down into the mediastinum behind the sternum and using gentle traction, which may be aided by the use of a bent dessert spoon when dealing with a very large multinodular gland. Care should be taken to avoid injury to the RLN, which may occasionally lie on the surface of the retrosternal component. A mediastinal split is seldom necessary.

Variations in technique

There are many variations to the above-described steps, which are useful in different situations. It may be appropriate to mobilize the lower pole first in some instances, especially if there is significant upper pole enlargement and extension towards the base of the skull. Early division of the isthmus is also a useful technique to aid in the mobilization of the lobe, especially in large goitres.

Complications of thyroid surgery

Thyroidectomy is a commonly performed and safe surgical procedure with a low morbidity and negligible mortality when performed by appropriately trained surgeons. General complications are those of anyone undergoing a general anaesthetic such as cardiac events, chest infection and venous thromboembolism, but these are uncommon following thyroid surgery with a current mortality rate in several large series approaching zero. The morbidity of thyroidectomy from its specific complications, however, continues to be a matter of concern (Box 18.2). Clearly, meticulous attention to operative technique is required and this is now an area for the trained endocrine surgeon rather than a general surgeon. Litigation for thyroidectomy complications amount to approximately 5% of general surgical claims, most of which involve RLN injury.

To avoid damage to the RLN a detailed knowledge of its variable anatomy and identification during surgery is essential. Bilateral palsy is exceedingly rare but may lead to temporary or permanent tracheostomy. This is most likely to be a problem in redo surgery when one RLN has already been permanently damaged. The frequency of RLN injury following thyroid surgery should be <1%, although audits of this outcome should reflect case mix and operative experience. The EBSLN is also at risk during thyroidectomy and permanent voice damage following its injury may occur and may be difficult to detect on indirect laryngoscopy. Such injury may be minimized if the nerve is identified and preserved during superior thyroid artery ligation.

Parathyroid damage producing hypocalcaemia is the second largest category of thyroid-related medicolegal claims and, although usually temporary, a long-term hypoparathyroid state has been shown to occur in up to 5% of cases. Most cases occur because of interruption of the arterial supply or obstruction of venous drainage, although inadvertent excision may also occur.

Hypothyroidism after total thyroidectomy is avoided by thyroxine replacement therapy, but can also occur with time after a subtotal resection. Recurrent hyperthyroidism after a subtotal resection presents more of a problem, as reoperation is associated with a significant increase in complications. Radioiodine ablation is probably a safer alternative to redo surgery.

BOX 18.2 Specific complications of thyroidectomy

- Respiratory distress: immediate postoperative period
  - RLN palsy
  - Acute laryngeal oedema
  - Wound haematoma
- Voice change (hoarseness, weakness, loss of high pitch): transient or long term
  - Injury to the EBSLN
  - RLN palsy
- Hypoparathyroidism: transient or long term
  - Bleeding
  - Infection
  - Hypothyroidism
  - Hypertrophic scarring or keloid

Postoperative reactionary haemorrhage is potentially catastrophic but can be avoided with meticulous haemostasis. However, the most serious and life-threatening complication is postoperative airway obstruction due to acute laryngeal oedema, which may or may not be associated with haematoma. The oedema is thought to be related to impaired lymphatic drainage of the larynx causing this internal oedema.

A well-positioned collar incision along a skin crease (if possible) gives adequate exposure and an excellent cosmetic result. However, wound complications occasionally occur and include a suture granuloma (which can be minimized by the use of absorbable suture material or clips), infection (cellulitis, abscess formation), hypertrophic scarring and keloid formation (especially in patients with an underlying predisposition).

Neck dissection

Radical neck dissection was first described by Jawdynski, a Polish surgeon in 1888. It was, however, popularized in the early twentieth century by Crile, a US surgeon, with whose name the operation remains synonymous today. The radical nature of the operation and the associated morbidity led to modifications of the technique in which vital structures such as the sternocleidomastoid, internal jugular vein and the spinal accessory nerve were preserved. These modifications led to the procedure being called 'modified radical' or 'functional' neck dissections and were developed by surgeons including Suarez, Bocca, Gavilan and Ballantyne. Further modifications involving excision of one or more lymph node groups in select compartments (levels) of the neck were termed 'selective' neck dissection.

Neck dissection terminology was standardized in 1991 by the American Academy's Committee for Head and Neck Surgery and Oncology. Lymph node groups were classified into levels, and six levels of the neck were initially described (level VII was added at a later stage). The levels in the neck are described below and depicted in Figure 18.15a,b.

- *Level I* is bounded by the midline medially, the body of the mandible superiorly, the posterior belly of the digastric laterally and the hyoid bone inferiorly. This level is further subdivided into level IA (the submental triangle) and level IB (the submandibular triangle) separated by the anterior belly of the digastric muscle.
- *Level II* is a triangular area bounded by the posterior belly of the digastric/stylohyoid muscle anteriorly, the posterior border of the sternomastoid posteriorly and an imaginary line along the inferior border of the hyoid bone inferiorly. The spinal accessory nerve divides this into a level IIA (anteroinferior to the nerve) and level IIB (posterosuperior to the nerve).
- *Level III* is the area below level II to the level of the cricothyroid notch (clinical landmark) or omohyoid (surgical landmark). The posterior border is again formed by the posterior edge of the sternocleidomastoid and the medial/anterior boundary is the lateral border of the sternohyoid muscle.
- *Level IV* is the area below level III, with the other boundaries being the posterior border of the sternocleidomastoid posteriorly, the clavicle inferiorly and the lateral border of the sternohyoid muscle anteromedially.

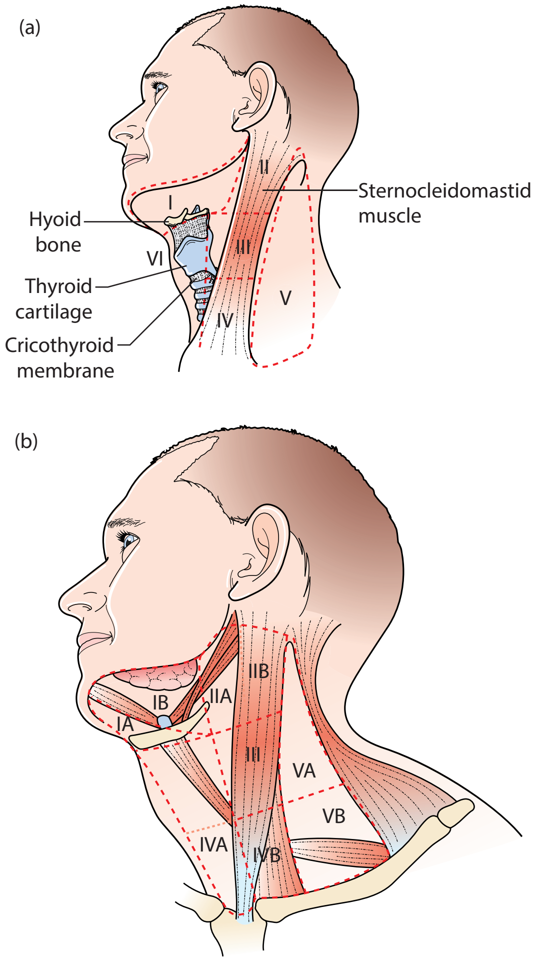


Figure 18.15 (a,b) Levels of the neck (see text for description).

- *Level V* is a triangular-shaped area bordered anteriorly by the posterior border of the sternocleidomastoid, the clavicle inferiorly and the anterior border of trapezius posteriorly. This is further subdivided into an upper VA around the spinal accessory nerve and a lower VB (or supraclavicular nodes) around the transverse cervical vessels; these being divided by the inferior belly of the omohyoid muscle.
- *Level VI* is also called the anterior compartment; this extends from the hyoid bone superiorly to the suprasternal notch inferiorly and the medial border of the carotid sheath laterally.
- *Level VII* is strictly mediastinal and refers to the area behind the manubrium and extends from the suprasternal notch superiorly to the brachiocephalic vein inferiorly.

General principles of neck dissection in thyroid cancer

Thyroid cancers are in general slow growing and lymph node metastases in thyroid cancer have not been shown to influence survival. Although berry picking of involved lymph nodes is to

Table 18.1 The vital structures in the various neck compartments (apart from the carotid sheath vessels, trachea and oesophagus) and consequences of their injury

	Vital structures	Consequence
Central compartment dissection	Parathyroid glands	Hypoparathyroidism
	Recurrent laryngeal nerves	Voice change, aspiration, tracheostomy (very rare)
	External branch of the superior laryngeal nerve	Inability to raise pitch of voice
Lateral compartment dissection	Salivary glands (level I) – very rare	Salivary fistula
	Marginal mandibular branch of the facial nerve (level I) – very rare	Drooping of the angle of the mouth on the ipsilateral side
	Spinal accessory nerve (levels II and V)	Drooping of the shoulder, and neck and shoulder dysfunction
	Hypoglossal nerve (level II)	Atrophy of the tongue and deviation towards the side of the palsy
	Vagus nerve	Vocal cord palsy (see as for the recurrent laryngeal nerve)
	Sympathetic chain	Horner syndrome (ptosis, miosis, anhidrosis, enophthalmos and loss of ciliospinal reflex)
	Cervical plexus injury	Numbness and paraesthesiae (often settles spontaneously)
	Thoracic duct (level V)	Chyle leak/fistula
	Phrenic nerve (level III/IV)	Ipsilateral diaphragmatic palsy
	Brachial plexus (level V)	Sensory and motor deficit affecting the shoulder and upper limbs

be condemned, the other extreme of radical and supraradical neck dissection is largely unnecessary and harmful in patients with thyroid cancer. Selective node dissection is the most commonly employed procedure.

The need for and the extent of lymph node dissection in thyroid cancer has already been discussed. Thyroid surgeons often refer to levels VI and VII as the central compartment and levels II–V as the lateral compartment. The commonest levels involved and operated upon in thyroid cancer are levels VI and VII (central compartment). Prophylactic central compartment neck dissection is routine in MTC and often selective in PTC. Lateral compartment lymph node dissection in thyroid cancer is usually done as a therapeutic procedure (i.e. when lymph nodes are palpable or shown to be pathological) and involves a selective clearance of levels IIA, III, IV and VB. Involvement of levels I, IIB and VA is very rare.

As mentioned previously in relation to thyroidectomy, a clear explanation of the need for and planned extent of lymphadenectomy should be given preoperatively. Complications specific to lymphadenectomy such as seroma formation, wound infection, bleeding from major vessels and damage to surrounding vital structures should be discussed. The vital structures in the various neck compartments (apart from the carotid sheath vessels, trachea and oesophagus) and consequences of their injury are shown in Table 18.1.

GUIDE TO FURTHER READING

Abraham P, Acharya S. Current and emerging treatment options for Graves' hyperthyroidism. *Ther Clin Risk Manag* 2010;**6**:29–40.

Cooper DS, Doherty GM, Haugen BR, *et al.* Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2009;**19**: 1167–1214.

Fast S, Nielsen VE, Bonnema SJ, Hegedüs L. Time to reconsider nonsurgical therapy of benign non-toxic multinodular goitre: focus on recombinant human TSH augmented radioiodine therapy. *Eur J Endocrinol* 2009;**160**:517–528.

Kuma K, Matsuzuka F, Kobayashi A, *et al.* Outcome of long standing solitary thyroid nodules. *World J Surg* 1992;**16**:583–587.

Kuma K, Matsuzuka F, Yokozawa T. Fate of untreated benign thyroid nodules: results of long term follow up. *World J Surg* 1994;**18**:495–498.

Nagaiah G, Hossain A, Mooney CJ, *et al.* Anaplastic thyroid cancer: a review of epidemiology, pathogenesis, and treatment. *J Oncol* 2011;**2011**:542358.

Werga P, Wallin G, Skoog L, *et al.* Expanding role of fine-needle aspiration cytology in thyroid diagnosis and management. *World J Surg* 2000;**24**:907–912.

CHAPTER 19

Disorders of the parathyroid glands

PAUL R. MADDOX

Parathyroid embryology, anatomy and physiology

Embryology

The superior parathyroid glands originate from the dorsal tips of the fourth pharyngeal pouch, which is incorporated into the lateral aspect of the thyroid along with the ultimobranchial body; this common origin occasionally leads to an intrathyroidal location for the superior parathyroid gland, although this is rare.

The inferior parathyroid glands arise from the dorsal aspect of the third pharyngeal pouch, with the thymus originating from the ventral aspect. Together, they descend as a complex in a plane ventral to the fourth pharyngeal pouch, and the lower parathyroids are therefore found in a more anterior position than the upper parathyroids, usually dissociating from the thymus near the lower pole of the thyroid, but this migration can vary widely. With an absence of migration, the inferior parathyroid gland may be found superior to the upper pole of the thyroid mimicking a superior gland, but surrounding thymic tissue clarifies the true origin. If the inferior parathyroid remains attached to thymus, it will migrate to the anterior mediastinum.

Surgical anatomy

Awareness of the common pathways of migration is invaluable in parathyroid surgery (Figure 19.1). Eighty per cent of the upper parathyroid glands lie within a localized area of a 2 cm radius, cranial to the intersection of the recurrent laryngeal nerve (RLN) and the inferior thyroid artery (ITA). This is a very common symmetrical position for the superior glands, which are usually tucked away posterior to the upper pole of the thyroid. If the glands are sited more anteriorly they are located on the surface of the thyroid frequently beneath its capsule, where there is typically freedom of movement as opposed to prominent thyroid nodules which are fixed. Approximately 1% of superior glands are found in the paraoesophageal or retro-oesophageal space, from where they may descend to the posterior mediastinum owing to the effect of negative intrathoracic pressure of respiration.

More than half of the inferior parathyroid glands are located around the lower pole of the thyroid, with a quarter being found within the thyrothymic ligament or within the thymus itself. As the inferior gland becomes enlarged, it tends to migrate into the thymus within the anterior mediastinum, where up to one-third of all missed parathyroid tumours can be found.

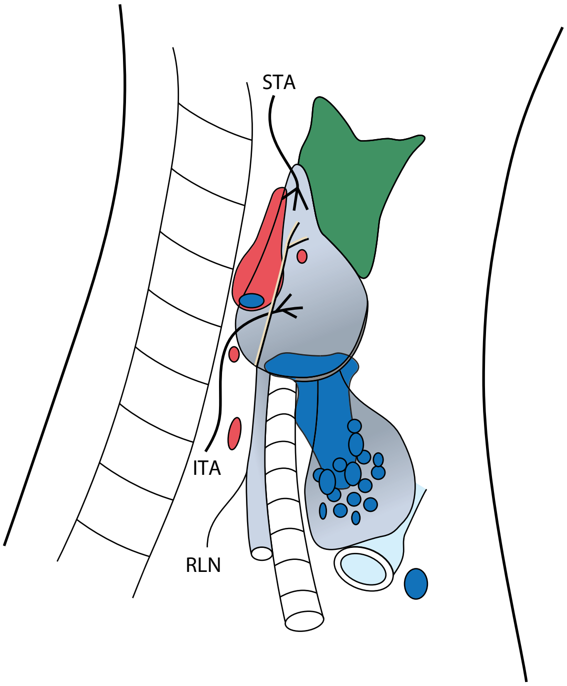


Figure 19.1 Sites of migration rest for superior (shaded red) and inferior (shaded blue) parathyroid glands. RLN, recurrent laryngeal nerve; STA/ITA, superior/inferior thyroid artery.

Parathyroid physiology

The parathyroid glands secrete parathormone (PTH). PTH is a polypeptide with 84 amino acid residues and a molecular weight of 9500. It is formed from preproPTH (115 amino acids), which is cleaved in the endoplasmic reticulum to form proPTH (90 amino acids). Six further amino acids are cleaved in the Golgi apparatus and the PTH is then stored in secretory granules in the chief cells of the parathyroid glands. The N-terminal portion is felt to be the biologically active portion of PTH. The half-life of PTH is only 10–15 minutes, as it is rapidly cleaved by the Kupffer cells in the liver. The main mechanism for the regulation of PTH is a negative feedback with plasma ionized calcium. Accurate measurement of circulating intact PTH is now obtained using two-site radioimmunoassay, with an assay against both the C-terminal and the N-terminal. This enables it to be differentiated from parathyroid hormone-related protein (PTHrP), which has PTH-like activity. PTHrP has a similar N-terminal, but has 140 amino acids, and is found in breast milk, the renal tubules and the brain. PTHrP is also responsible for 80% of the hypercalcaemia associated with malignancy – the humoral hypercalcaemia of malignancy.

PTH has effects on the bone, kidney and the gastrointestinal tract. In the bone, it promotes bone reabsorption, by the activation of osteoblasts to mobilize calcium. In the kidney, PTH increases the excretion of phosphate, by decreasing proximal tubular reabsorption. In normal circumstances, PTH increases calcium reabsorption at the distal tubules. However, this effect is overwhelmed in patients with hyperparathyroidism (HPT) by the raised levels of calcium being filtered, leading to hypercalciuria. This effect helps distinguish primary HPT from rarer causes of hypercalcaemia, such as benign familial hypocalciuric hypercalcaemia (FHH). PTH also converts 25-hydroxycholecalciferol to the active 1,25-hydroxycholecalciferol, which aids the mobilization of calcium from the bone and promotes the absorption of calcium from the gastrointestinal tract.

Hyperparathyroidism

Primary HPT is the inappropriate secretion of PTH associated with excessive growth and abnormal function of one or more parathyroid glands. These abnormalities cause hypercalcaemia in most but not all patients ('normocalcaemic' HPT). The prevalence of HPT is 0.1–2%, being highest in postmenopausal women. HPT occurs most commonly in its sporadic form, familial cases accounting for <5% of cases.

Diagnosis

Although patients with HPT may present with severe classic symptoms and signs, these are relatively uncommon today in the UK. The most common complaints are non-specific and are vague (Box 19.1). The symptoms of HPT do not correlate with the severity of the hypercalcaemia; many patients (at least 30%) are diagnosed consequent to the incidental detection of a raised serum calcium. Although these patients are often described as having asymptomatic HPT, the term mild HPT is more appropriate because truly asymptomatic disease is uncommon. The prevalence of neuropsychiatric symptoms and cardiovascular risk factors in such individuals is frequently ignored.

BOX 19.1 Symptoms and signs of hyperparathyroidism

- Renal
  - Urinary frequency/polyuria and thirst
  - Urinary tract calculi
  - Nephrocalcinosis
  - Renal failure
- Bone
  - Bone and joint pains
  - Chondrocalcinosis
  - Reduced bone mineral density
  - Fracture
- Gastrointestinal
  - Poor appetite
  - Nausea and vomiting
  - Constipation
  - Abdominal pain
  - Peptic ulcer
  - Pancreatitis
- Cardiovascular
  - Hypertension
  - Various myocardial changes
- Neuropsychiatric
  - Depression, psychosis
  - Fatigue, weakness, poor memory

The diagnosis of HPT is confirmed by an inappropriately high level of intact PTH – the biologically active hormone – in a patient with hypercalcaemia. If the PTH level is low normal or suppressed, the hypercalcaemia is not caused by HPT. The differential diagnosis of HPT (Box 19.2) includes hypercalcaemia of malignancy mediated by PTHrP and other growth factors secreted by the primary tumour that stimulate osteoclastic bone resorption and the presence of multiple bone metastases. Patients with FHH should be excluded by the absence of symptoms, a family history of the disorder and/or a failed neck exploration, low urinary calcium and a calcium creatinine clearance ratio <0.01.

Pathology

At least 80% of patients with HPT will have single gland disease – a parathyroid adenoma. Most abnormal glands will weigh 200–1000 mg (normal glands weighing up to 50 mg). They are usually reddish brown in colour and have a low fat content, in contrast to a suppressed normal gland. A remnant of normal compressed tissue may be visible to the pathologist; this 'rim' of parathyroid may be used as a criterion to distinguish parathyroid adenoma from hyperplasia. Multiple gland disease in HPT may be due to multiple adenomas (more frequent in older patients) in up to 10% of patients or parathyroid hyperplasia (sporadic or associated with familial disease). In an individual patient with parathyroid hyperplasia the glands may be of varied size.

BOX 19.2 Causes of hypercalcaemia

- Primary hyperparathyroidism
  - Primary hyperparathyroidism
  - Familial hyperparathyroidism
- Malignancy
- Sarcoidosis
- Vitamin D excess
- Thyrotoxicosis
- Drugs – thiazide diuretics, lithium
- Familial hypocalciuric hypercalcaemia
- Prolonged immobilization

Two histological subtypes are recognized – chief cell hyperplasia and water clear-cell hyperplasia.

Preoperative localization studies

An experienced parathyroid surgeon will cure at least 95% of patients at initial bilateral neck exploration without the need for preoperative localization. On that basis, the routine use of localization studies has not been previously recommended. The use of unilateral neck exploration and the arrival of minimally invasive parathyroid surgery have, particularly in European and US institutions, led to renewed interest in scan-directed first-time surgery with a satisfactory outcome/cost-benefit.

The most frequently used non-invasive localization studies are:

- *Ultrasound*. Published series report that 50–80% of enlarged parathyroid glands can be identified prior to surgery. The skill of the user, the size and position of the abnormal gland and the presence of coexisting thyroid nodules will affect the sensitivity, specificity and accuracy of this technique.
- *Technetium-99m sestamibi (MIBI).* Up to 85% of parathyroid adenomas and 50% of abnormal glands in patients with multiglandular disease can be localized. The use of single photon emission CT (SPECT) allows sagittal and transverse scans and better localization, particularly of ectopic parathyroid tissue.

Therefore, patients with an accurate and reliable localization (concordant scans) of single gland disease (approximately 85%) may be treated with a focused minimally invasive approach.

The use of CT, MRI, selective angiography, selective venous catheterization for PTH gradients and positron emission tomography scans is well described, but these are best reserved for localization of abnormal parathyroid glands when initial surgery has failed.

Natural history of primary hyperparathyroidism

There is no doubt that patients with significant hypercalcaemia and or symptomatic HPT should undergo neck exploration and excision of abnormal parathyroid glands. There is longstanding dispute as to what is appropriate management for patients with minimally elevated serum calcium levels and/or minimal symptoms. In this group, surgery results in a reduction in the HPT-associated risk of death from cardiovascular disease, an increase in cancellous bone density, reversal of neuromuscular deficit and improvement in quality of life. Non-surgical treatment of HPT consists of observation and long-term follow-up. Hormone replacement therapy for postmenopausal women with HPT at risk of osteoporosis does reduce the serum calcium. Although long-term observational follow-up studies show that as many as 25% of untreated patients have progression of their disease, there are no prospective randomized controlled trials on which to base guidelines for treatment. On that basis, a liberal approach to surgery is recommended.

Familial hyperparathyroidism

Multiple endocrine neoplasia type 1 (MEN 1)

An autosomal dominant disorder, the causative gene is located on chromosome 11q13, and is characterized by HPT (90%), multiple neuroendocrine tumours of the pancreas (75%; usually gastrinoma) and pituitary tumours (>30%; usually prolactinoma). The HPT is multiglandular. A surgical intervention less than subtotal parathyroidectomy is associated with recurrence rates of up to 40%. Supernumerary parathyroid glands are common in this disease (up to 20%), most frequently sited in the thymus. The surgical approach should therefore include *at least* subtotal parathyroidectomy and transcervical thymectomy.

Multiple endocrine neoplasia type 2 (MEN 2A and MEN 2B)

An autosomal dominant disorder, the gene is localized to chromosome 10q, and is characterized by medullary thyroid cancer (invariable), phaeochromocytoma (50%) and HPT (25%). Patients with MEN 2B have a Marfanoid phenotype and multiple mucosal ganglioneuromata. The HPT of MEN 2 is less severe; single gland disease is more frequent; and only enlarged parathyroid glands should be removed at operation.

Familial non-MEN HPT is very rare and frequently associated with multiglandular disease, supernumerary glands and parathyroid carcinoma. In some families the disorder is associated with mandibular tumours.

It is recommended that patients and families with genetically determined disease should be treated in specialist centres, as the treatment and screening strategies required for both MEN 1 and MEN 2 require a multidisciplinary approach that includes endocrine surgeons and physicians, clinical and molecular geneticists, interventional radiologists, histopathologists, biochemists and pathologists.

Hypercalcaemic crisis in hyperparathyroidism

Acute severe hypercalcaemia (corrected calcium >3.5 mmol/L) may occur in patients with HPT, often precipitated by intercurrent illness (associated with vomiting/diarrhoea) or diuretic therapy. The clinical presentation may include fatigue, weakness, polyuria, confusion, coma, pancreatitis and cardiovascular instability. Hypotension, pancreatitis and cardiac failure predict a poor outcome.

Treatment

This consists of:

- Aggressive rehydration with normal saline (up to 200 mL/h) to promote a good urine output. This may require monitoring of central venous pressure in elderly or unfit patients. Diuretic therapy is best avoided.
- Bisphosphonates act by inhibiting bone resorption. Treatment is usually by giving pamidronic acid 60–90 mg intravenously over 4–24 hours. Multiple doses may be required.

Surgery should be delayed until the corrected serum calcium is below 3 mmol/L.

Parathyroid carcinoma

The incidence of parathyroid carcinoma is approximately 1% of primary HPT cases. Patients will typically have a long history, calcium levels >3.5 mmol/L (70%), a palpable neck mass (50%) and severe bone disease (60%). At operation, the diagnosis is made on the basis of a firm, adherent grey–white gland (sometimes an atypical 'old' benign adenoma) or local invasion. In this situation:

- biopsy is contraindicated
- *en bloc* resection of the parathyroid gland, thyroid lobe and isthmus and other involved tissue should be performed.

Completion surgery should be performed if the diagnosis is made by the pathologist (a thick fibrous capsule with trabeculated fibrous bands traversing the tumour). Half the patients will have a favourable prognosis, although recurrent disease (local/regional and/or metastases with hypercalcaemia) usually occurs within 3 years.

Surgical approaches to hyperparathyroidism

The definitive treatment for patients with primary HPT is surgery, which should ensure a high rate of cure with a low complication rate. Doppman noted that 'the only localizing study indicated in a patient with untreated primary HPT is to localize an experienced parathyroid surgeon'. It is therefore axiomatic that the surgeon should have a good understanding of embryology and surgical anatomy of the parathyroid glands to ensure a successful outcome (see Parathyroid embryology, anatomy and physiology).

Pathology

The normal shape of a parathyroid gland depends on its location. In loose tissue it classically appears as a light-brown ovoid or sphere floating within surrounding yellow fat, but under a capsule the shape is flat with sharp edges. When enlarged, however, it assumes a spherical shape, indicating hyperactivity of the parathyroid gland. The greatest diameter of a normal parathyroid gland is usually no more than 5 mm with an upper limit of normal weight of 50 mg.

Surgical technique

Unilateral or bilateral exploration

Unilateral neck exploration for primary HPT is controversial because of the concern about missing multiglandular disease and the inaccuracy of preoperative localization tests. Protagonists claim decreased complications and cost-effectiveness for the technique but drawbacks are obviously missing double adenomas or hyperplasia. Retrospective data analysis reveals no differences between the two techniques and to date no prospective studies have compared unilateral with bilateral exploration. Bilateral neck exploration has, therefore, previously been the standard approach because it is safe, avoids missing a second adenoma and avoids unnecessary expensive preoperative localization studies as >95% of patients can be cured when treated by an experienced endocrine surgeon.

However, over the last decade, popularity for a focused unilateral technique has increased and with concordant localization a minimally invasive parathyroidectomy has now become the standard approach for putatively single gland disease.

Cervical exploration for primary hyperparathyroidism

Preoperative preparation is as previously described for thyroidectomy except for the use of methylene blue (5 mg/kg body weight methylene blue in 500 mL dextrose/saline infused intravenously over 1 hour before surgery), which selectively stains parathyroid tissue and is a valuable aid to the difficult case. It is fundamental for the endocrine surgeon to identify all four (or possibly more) parathyroid glands in the neck. Through a collar (skin crease) incision the parathyroid exploration is usually started by searching for the upper parathyroid gland just above the neurovascular intersection of the RLN/ITA on the dorsum of the thyroid gland, retracting the thyroid with a finger on a gauze swab. Careful dissection of the fat and fibrous attachments to the thyroid at this point usually reveals the majority of upper glands. The dissection should be under direct vision at all times to prevent injury to the RLN, which usually runs anteromedially to the upper gland. When it cannot be found in its usual site exploration of the para/retro-oesophageal space is undertaken with digital exploration of the posterior mediastinum and close examination of the upper thyroid pole to inspect for a subcapsular intrathyroidal gland.

The search for the inferior parathyroid gland begins with a thorough inspection of the lower thyroid pole, the thyrothymic ligament and the thymus itself. Usually, the lower parathyroid lies in the fat between the inferior thyroid veins, but rarely it may be found in the anterior surface of the thyroid gland.

Bilateral exploration should be undertaken and in 85% of patients there will be a solitary parathyroid adenoma. The adenoma is gently dissected using a gauze pledget (to push the gland forward) and fine clips (taking care not to break the capsule) to release it from its fascial coverings. The pedicle is defined, clipped and divided, and the adenoma removed (Figures 19.2 and 19.3).

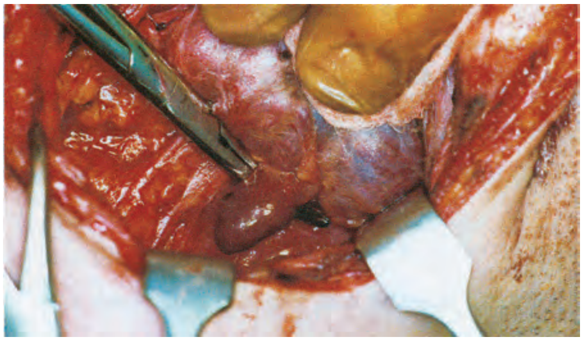


Figure 19.2 Cervical exploration. Left superior parathyroid adenoma (stained with methylene blue) being gently 'teased out' from surrounding fascia.

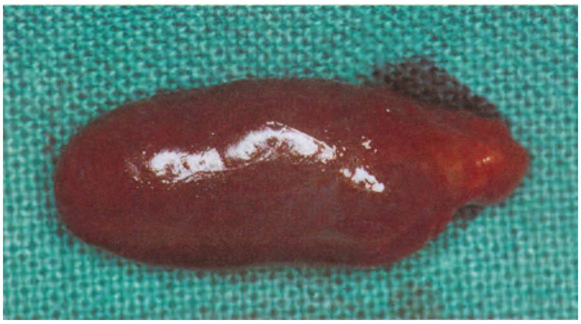


Figure 19.3 Left superior parathyroid adenoma weighing 630 mg; note the intact glistening capsule.

If three normal parathyroid glands are found the adenoma should be removed and examined by frozen section. It is controversial whether a biopsy of one of the normal glands is undertaken but the author does not routinely biopsy unless the appearances at operation are suspicious. If two enlarged glands are found they should be removed and the two normal glands should be biopsied and marked with clips. Microscopic distinction between adenomas and hyperplasia can be difficult and the role of the pathologist intraoperatively is therefore limited to the identification of parathyroid tissue. If all parathyroid glands are enlarged three parathyroid glands should be removed along with a half of the fourth and a thymectomy should be performed, as supernumerary parathyroid glands are frequently located in the thymus. The parathyroid remnant should be approximately 50 mg of tissue and an easily accessible gland with a reliable vascular stalk should be chosen and biopsied first before the others are removed to ensure that it remains viable. An alternative surgical strategy is to remove all four glands and reimplant half of one as an autotransplant into a pocket of the sternomastoid muscle.

Minimally invasive parathyroid surgery

This has now become the standard approach for unilateral disease selected by concordant scans for approximately 80% of cases. Focused surgery is performed through a small (approximately 2 cm) lateral skin crease directly over the site of the abnormal gland. The scar fades to a barely visible line over a few months. This procedure can be performed under local anaesthetic, cervical block or laryngeal mask general anaesthesia, usually in less than 30 minutes and as a day-case procedure. Conversion to a full cervical exploration may be necessary but is easily facilitated by extending the wound along a standard collar incision (approximately 5 cm) but would require a general anaesthetic.

Intraoperative PTH monitoring (with rapid assay) has been advocated as evidence of biochemical cure. Normal PTH levels occur within 10 minutes of removal of the abnormal gland. However, this is an expensive and potentially time-wasting procedure and the author currently prefers to rely on frozen section alone.

Minimally invasive parathyroid surgery is reliable and safe and has now become the first-line treatment for primary HPT with focused imaging from concordant scans.

Secondary hyperparathyroidism

In patients with secondary HPT, familial HPT or MEN 1 syndrome, all parathyroid glands are involved, and the rate of recurrent HPT in these patients ranges from 10% to 15%. A subtotal parathyroidectomy should therefore be performed, leaving behind a remnant of a lower parathyroid (with a clip marker) or autotransplantation is performed as described above. Cryopreservation of some parathyroid tissue may also be considered.

Mediastinotomy

Following a thorough exploration of the neck, if an abnormal gland is not found, normal glands should never be removed and the cervical wound should be closed and full investigation carried out; including high-resolution ultrasound scanning, CT, MRI, sestamibi subtraction scan and selective venous sampling to confirm the location of the ectopic adenoma. This is usually found within the chest (Figure 19.4) in either the anterior mediastinum (for an inferior gland) or posterior mediastinum (for a superior gland). Surgery can then be planned accordingly for either a mediastinotomy or lateral thoracotomy, respectively (Figure 19.5) when inaccessible from the neck. More recently, thoracoscopy or mediastinoscopy has

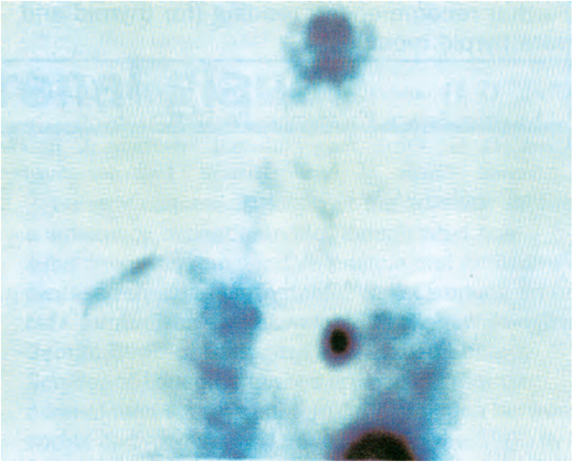


Figure 19.4 Sestamibi scan demonstrating ectopic adenoma of a fifth parathyroid gland in the aortopulmonary window of the middle mediastinum.

been reported to be a successful minimally invasive technique to remove mediastinal adenomas.

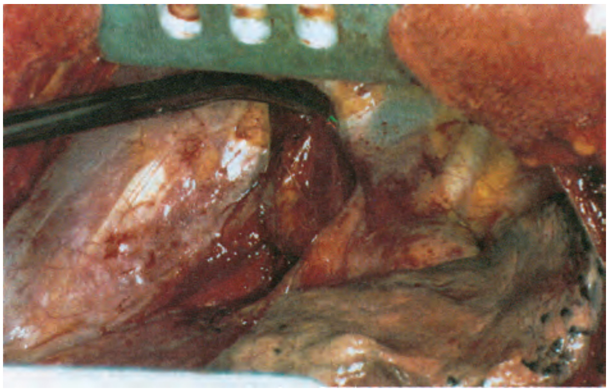


Figure 19.5 Removal of an ectopic fifth parathyroid adenoma (7 g) from the aortopulmonary window through a left fourth rib thoracotomy. (Courtesy of Mr C. Forrester-Wood, Bristol Royal Infirmary, UK.)

GUIDE TO FURTHER READING

Clark OH, Duh Q-Y (eds). *Textbook of Endocrine Surgery*. Philadelphia, PA: W.B. Saunders, 2005.

Clark OH, Duh Q-Y, Siperstein AE. *Endocrine Surgery*. Philadelphia, PA: W.B. Saunders, 1995.

Frieser SR, Thompson NW (eds). *Surgical Endocrinology: Clinical Syndromes*. Philadelphia, PA: J.B. Lippincott, 1990.

Lynn J, Bloom SR (eds). *Surgical Endocrinology*. Oxford, UK: Butterworth-Heinemann, 1993.

Wheeler MH, Richards SH. The laryngeal nerves. *Curr Surg Practice* **5**:237–250, 1993.

CHAPTER 20

Disorders of the adrenal glands

GREG P. SADLER

Introduction

The Roman anatomist Bartholomaeus Eustachius first described the adrenal glands in 1552 in his *Opuscula Anatomica*, referring to them as 'glandulae renibus incumbentes' (glands lying on the kidneys). The function of the adrenal glands, however, remained a mystery to workers for 300 years, until in 1856 the French physiologist Charles Brown-Séquard demonstrated that these glands were essential to life in a series of experiments on dogs and rats. Brown-Séquard had been inspired by a report the previous year by Thomas Addison of London describing the clinical findings of 11 patients in whom the adrenal glands had been found at postmortem to be destroyed, through either tuberculosis or cancer (a condition later termed Addison disease).

General consensus held that the adrenals secreted a substance essential to life, though what this substance was remained an enigma and Addison disease remained untreatable. While working on a cure for Addison disease in 1893 the physiologists George Oliver of Harrogate and Edward Sharpey-Schäfer of London prepared an extract from the adrenal medulla (the anatomical distinction between cortex and medulla had been made in 1805 by Baron Georges Cuvier and the terms coined by Emil Huschke in 1845). The extract failed to cure Addison disease but produced a marked constriction of blood vessels. This preparation was later purified and termed adrenaline (1897) or epinephrine (1901). It was not until the late 1920s when US workers prepared an adrenal cortical extract termed 'cortin' that Addison disease was treated successfully.

The adrenal glands are now known to secrete a wide variety of hormones that regulate many essential physiological functions including blood pressure, carbohydrate, fat and protein metabolism. It is not surprising, therefore, that development of hyperplastic or neoplastic conditions may lead to overproduction of various hormones, resulting in a wide variety of clinical syndromes.

The current-day endocrine surgeon with an ever-increasing diversity of biochemical tests, pharmaceutical agents, localizing modalities and operative techniques has a unique opportunity to correct profound physiological and often life-threatening disturbances and restore the patient to complete normality.

Anatomy

The adrenals are retroperitoneal structures, each weighing approximately 2–6 g and located on either side of the vertebral column within Gerota's fascia, in close proximity to the crura of the diaphragm and the superior poles of the kidneys. The right adrenal gland is triangular or pyramidal in shape with the apex superiorly and the base towards the right kidney. It lies on the right crux of the diaphragm posteriorly, the bare area of the liver anteriorly, the inferior vena cava anteromedially and the superior pole of the right kidney inferolaterally. The left gland is crescentic in shape, related anteriorly to the stomach and pancreas and posteriorly to the diaphragm. Anteriorly the gland is covered by the peritoneum of the lesser sac. Arterial supply is drawn from the aorta, renal artery and inferior phrenic artery, though some tumours (particularly phaeochromocytoma) may demonstrate marked vascular neogenesis. Main venous drainage on the right side is via a short adrenal vein (5 mm) directly to the inferior vena cava, usually entering posteriorly. On the left side drainage is directly to the left renal vein. Frequently venous drainage is supplemented by additional veins running directly to the vena cava or hepatic veins on the right and the phrenic vein on the left.

Macroscopically the adrenal glands are a golden yellow colour, which contrasts with the surrounding perinephric fat. When sectioned the outer cortical layer (which makes up 80% of the normal gland in volume) is lipid rich and golden or yellow, in contrast to the well-vascularized medulla, which tends to be reddish-brown in colour. The cortex consists of three layers. From outer to inner these are the zona glomerulosa (aldosterone production), zona fasciculata (cortisol production) and zona reticularis (androgen production). The adrenals have a rich sympathetic nerve supply from the coeliac plexus but preganglionic fibres from the splanchnic nerves pass through the coeliac plexus to supply the adrenal medulla.

Physiology

The principal hormone secretions of the adrenal gland – aldosterone, cortisol and the androgens (mainly testosterone) – are synthesized via intermediates, such as pregnenolone and progesterone, from cholesterol by a series of metabolic steps, each requiring a specific enzyme.

Aldosterone is a mineralocorticoid with the principal action of stimulating sodium reabsorption in exchange for potassium and hydrogen ions. This occurs in the distal renal tubule, ascending loop of Henle and collecting ducts. Plasma oncotic pressure is increased, thus expanding plasma volume. Synthesis and secretion of aldosterone is regulated by angiotensin II and by plasma levels of sodium and potassium. Aldosterone is produced in the zona glomerulosa in response to a fall in blood volume, registered in the juxtaglomerular apparatus of the afferent tubules of the kidney. These cells release renin, which splits angiotensinogen to angiotensin I, which is then converted to angiotensin II by angiotensin-converting enzyme (ACE). Sodium retention is directly linked to potassium exchange, the rate being dependent on the rate at which sodium is presented to the tubule.

Glucocorticoids, mainly cortisol and small amounts of corticosterone, are produced in the zona fasciculata. The rate of production is directly dependent on adrenocorticotrophic hormone (ACTH) from the pituitary, which in turn is suppressed directly by cortisol, via a direct negative feedback mechanism of cortisol on the pituitary. ACTH secretion is stimulated by corticotrophin-releasing hormone (CRH), produced by the hypothalamus (CRH production is also suppressed by cortisol). Cortisol has a wide range of physiological actions involving fat, carbohydrate and protein metabolism and is subject to a specific diurnal variation in synthesis and production, with highest levels being achieved early in the morning and lowest levels at night.

Androgen production is from the zona reticularis and is also stimulated by ACTH. In the adult, however, the main site of sex hormone production is the gonads. Sex hormones produced from the adrenals play a role in growth and sexual differentiation.

Addison disease and adrenal insufficiency

Adrenal insufficiency was first recognized and described by Thomas Addison in a presentation in 1849. His classic 1855 publication documented the symptoms and physical findings in 11 patients with adrenal destruction noted at postmortem. The major causative factor was tuberculosis, although metastatic carcinoma and atrophy were also documented. The term Addison disease was first used by Armand Trousseau of Paris the following year. Trosseau had read Addison's monograph *On the Constitutional and Local Effects of Disease of the Supra-renal Capsules* and noted the same symptoms, clinical and postmortem findings in a patient with tuberculosis of the adrenal glands.

Addison disease has always been a rare condition and, until recently, a reduction in the numbers of patients with tuberculosis decreased the incidence even further, but the incidence of tuberculosis has risen and this, coupled with AIDS and autoimmune adrenalitis, has meant that the incidence of Addison disease has increased. In 1974 the recorded incidence in Denmark was six cases per 100 000 population.

A number of conditions have now been described that can destroy or interfere with the adrenal cortex, resulting in Addison disease. The most common cause is autoimmune disease, which is responsible for 65% of all cases of primary adrenal insufficiency, and association with other autoimmune conditions such as autoimmune thyroiditis and insulin-dependent diabetes is common. Other causes include tuberculosis, AIDS, sarcoidosis, amyloidosis, haemochromatosis, histoplasmosis, metastatic carcinoma, congenital adrenal hyperplasia, adrenocortical haemorrhage (particularly in anticoagulated patients), adrenal venography or meningococcal septicaemia (Waterhouse–Friderichsen syndrome).

AIDS is increasingly responsible for patients with Addison disease by a number of mechanisms. Adrenal destruction and insufficiency can be caused by infection with *Mycobacterium tuberculosis, Mycobacterium avium-intracellulare*, toxoplasmosis, histoplasmosis and *Pneumocystis carinii*. Neoplastic destruction by Kaposi's sarcoma or lymphoma has also been described. Ketoconazole, rifampicin, phenytoin and corticosteroids, drugs used in the treatment of AIDS-related illness, may all cause adrenal impairment. As patients with AIDS frequently present with similar symptoms to those with Addison disease, a high index of suspicion is required.

Bilateral surgical removal is an obvious cause of adrenal insufficiency. By definition, however, adrenal insufficiency following bilateral adrenalectomy is not Addison disease. Acute adrenal failure may also result following the sudden withdrawal of long-term steroid treatment, precipitating cardiovascular collapse or even death.

Diagnosis

Adrenal insufficiency is characterized by lassitude, weakness, anorexia and amenorrhoea. Physical findings include hypotension, pigmentation of the buccal mucosa (particularly on pressure areas), vitiligo and loss of body hair. In acute and extreme cases drowsiness, confusion, coma and death occur. In the collapsed patient, stigmata of disease for which steroids may be prescribed or bilateral adrenalectomy scars may provide important clues to the diagnosis. Features of adrenal, hypothalamic or pituitary disease may also be present.

The diagnosis of Addison disease may be confirmed by intramuscular injection of 250 μg of tetracosactrin and the plasma cortisol response at 0, 30 and 60 minutes recorded (short-term Synacthen test). In normal patients the plasma cortisol level should rise. The response is absent or impaired in patients with Addison disease. A long-term Synacthen test (depot test) helps to distinguish primary adrenal insufficiency from secondary adrenal dysfunction, usually from pituitary disease. Subsequent investigations including chest and abdominal radiographs and autoantibody screening are directed at establishing the cause of adrenal insufficiency.

Treatment of acute adrenal insufficiency

The signs and symptoms of acute adrenal insufficiency listed above progress rapidly. In patients with meningococcal septicaemia (often children) characteristic petechial skin haemorrhages may be noted in addition to hyperpyrexia, rigors and vomiting. The onset of profound shock with tachycardia and hypotension may occur suddenly. If appropriate therapy is not immediately instituted, coma and death ensue. The most common finding on routine laboratory testing is hyponatraemia, although this is not always present.

Once the diagnosis is suspected patients should be given 100 mg of hydrocortisone intravenously and rapidly resuscitated with intravenous normal saline. Intravenous antibiotics are given immediately if septicaemia is suspected as the underlying cause. Recovery is usually prompt but hydrocortisone (100 mg i.v.) should be continued every 6 hours until the patient is stabilized. Subsequently, patients can be changed to oral hydrocortisone or dexamethasone. Mineralocorticoids are not necessary when the dose of cortisol exceeds 60 mg/day. Chronic adrenal insufficiency will require maintenance therapy with hydrocortisone, usually 20 mg in the morning and 10 mg in the evening. Additional mineralocorticoid may be necessary in patients with impaired salt-retaining ability (0.05–0.10 mg fludrocortisone). Hypertension, oedema and hypokalaemia are signs of overtreatment. Fatigue, hypotension and hyperkalaemia are signs of undertreatment. The relative potencies of various steroid preparations are shown in Figure 20.1.

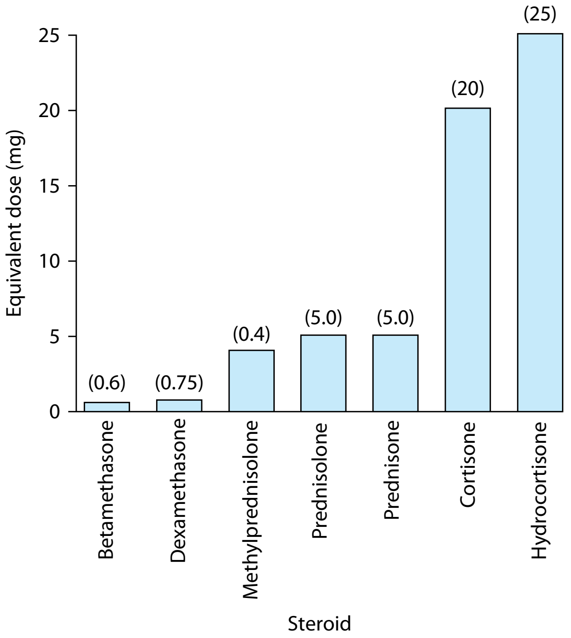


Figure 20.1 Relative potency of steroid compounds.

Surgery and steroid replacement therapy

- *Bilateral adrenalectomy.* Hydrocortisone sodium succinate is given i.m. (or i.v.) with the premedication and a further 100 mg given at the time of gland removal. In the postoperative period hydrocortisone 100 mg 6 hourly i.v. is continued for between 48 and 72 hours, reducing to 50 mg 6 hourly before the introduction of 20 mg of oral hydrocortisone in the morning and 10 mg at night. Mineralocorticoid is given as oral 0.05–0.1 mg fludrocortisone acetate daily.
- *Unilateral adrenalectomy* normally does not require steroid cover but, if performed for Cushing syndrome because of an adrenal adenoma, hydrocortisone sodium succinate 100 mg i.m./i.v. is given with the premedication and a further 100 mg given during the operation and continued 6 hourly, reducing with the introduction of 20 mg of oral hydrocortisone in the morning and 10 mg at night. Gradual reduction of steroid replacement is made on a monthly basis with complete weaning often taking up to 12–18 months, as the remaining adrenal gland recovers normal function.
- *Surgery in the hypoadrenal patient*, e.g. after bilateral adrenalectomy and in patients with adrenal suppression from steroid therapy. If these patients are having major surgery, additional steroids are required: hydrocortisone sodium succinate 100 mg i.m./i.v. is given with the premedication and a further 100 mg i.v. given during the operation and continued 6 hourly until the patient is able to return to the usual oral replacement steroid dose. If the postoperative period is prolonged, e.g. after a gastrointestinal operation, the parenteral hydrocortisone is reduced to 50 mg and then 25 mg 8 hourly until the patient is able to take oral medication.

For most minor surgery, in which oral intake is not restricted, intramuscular hydrocortisone is only required for 24 hours postoperatively and for minimal surgical procedures, such as endoscopy, a single dose of hydrocortisone 100 mg is all that is required. Patients who are on steroid therapy, who have undergone bilateral adrenalectomy or who suffer from primary hypoadrenalism must be fully advised by the medical staff on the nature of their condition and the symptoms to be expected with inadequate steroid replacement. They are instructed to double the usual dose of steroid immediately and seek medical assistance in the event of illness or severe stress. The crucial importance of this treatment must be emphasized. All patients should carry a steroid warning card and the wearing of a Medic Alert bracelet is to be encouraged. Patients who have had total adrenalectomy in some hospitals in the USA are given a steroid-containing syringe for emergency self-administration.

Congenital adrenal hyperplasia (adrenogenital syndrome)

The principal hormone secretions of the adrenal gland – aldosterone, cortisol and the androgens (mainly testosterone) – are synthesized via intermediates, such as pregnenolone and progesterone, from cholesterol by a series of metabolic steps, each requiring a specific enzyme. A deficiency in one of the enzymes (inherited in an autosomal recessive manner) leads to a deficiency in cortisol and/or aldosterone production. The lack of negative feedback to the pituitary results in excessive ACTH production. The growth-stimulating effect of ACTH leads to the development of bilateral adrenal hyperplasia.

The most common cause of adrenal hyperplasia is a deficiency of the enzyme 21-hydroxylase (90% of cases). This enzyme catalyses progesterone to deoxycorticosterone and 17-OH progesterone to 11-deoxycortisol. Deficiency leads directly to lack of production of aldosterone and cortisol. The backlog of metabolites is subsequently diverted into production of the androgens dehydroepiandrosterone, androstenedione and testosterone. Other enzyme deficiencies reported include 11 -hydroxylase, 3 -hydroxy steroid dehydrogenase and 17 -hydroxylase. Impaired production of aldosterone leads to *salt wasting* from excessive urinary salt loss. This is characterized in the newborn by vomiting and failure to feed and thrive.

The condition termed *congenital adrenal hyperplasia* is fortunately rare, with a reported incidence of one in 5000 to one in 15 000. Excessive androgen production in the female fetus leads to virilization, pseudohermaphroditism with ambiguous external genitalia, clitoral enlargement and fusion of the genital folds. In the male fetus the effects of androgen excess are not so obvious but penile enlargement occurs during childhood and pseudoprecocious puberty may ensue (infant Hercules). Early epiphyseal closure leads to short stature in adult life.

Diagnosis

Lack of the enzyme 21-hydroxylase leads to excess of the precursors progesterone and 17-OH progesterone. Measurement of the plasma level of 17-OH progesterone by radioimmunoassay is the most reliable method of confirming the diagnosis.

Treatment

Hydrocortisone (5 mg/day in divided doses) is administered to replace the glucocorticoid deficiency and totally suppress ACTH drive. Adequate suppression can be monitored by assessing 17-OH progesterone levels in urine and plasma. *Salt wasting* is treated by intravenous normal saline and mineralocorticoid replacement. Early diagnosis, adequate medical therapy and careful follow-up should ensure that growth development, fertility and life expectancy are all normal.

Cushing syndrome

The physiological and clinical features of Cushing syndrome are due to excessive levels of circulating glucocorticoids. The predominant clinical features of the disease include moon-type facies, truncal obesity with prominent interscapular fat pad formation, abdominal striae, diabetes and hypertension. The causes of Cushing syndrome are divided into two main groups: ACTH dependent and non-ACTH dependent.

Adrenocorticotrophic hormone-dependent Cushing syndrome

Cushing disease

Harvey Cushing (a Boston, MA, neurosurgeon) described the classic clinical features of the syndrome in 1932. He termed the condition 'pituitary basophilism', indicating that the condition was caused specifically by a pituitary basophil adenoma. Pituitary (ACTH)-dependent Cushing syndrome (referred to as Cushing disease) accounts for 70% of all cases of the syndrome. An ACTH-producing pituitary adenoma or microadenoma is responsible in approximately 75% of patients. Tumours are small (usually <1 cm) and malignancy is rare. The primary defect is probably located in the hypothalamus with overproduction of CRH. Chronically raised ACTH levels result in adrenal cortical hyperplasia and cortisol excess. ACTH secretion is still subject to negative feedback but at a higher set point than normal. Women are affected eight times more commonly than men and the disease is most common between 35 and 50 years of age.

Ectopic adrenocorticotrophic hormone syndrome

Ectopic ACTH production accounts for 10% of cases of Cushing syndrome and is caused by a spectrum of tumours originating from neuroendocrine cells. The commonest cause is a small cell carcinoma of the lung (previously referred to as oat cell tumours). Other causes include thymoma, medullary thyroid cancer, phaeochromocytoma, pancreatic islet cell tumours, certain ovarian cancers, pancreas, stomach and bronchial carcinoids.

The first case of ectopic ACTH syndrome was described by Brown in 1928, although the term 'ectopic ACTH syndrome' was originally coined by Liddle and colleagues in 1962. Cases caused by benign carcinoids may be difficult to diagnose as the onset of Cushing syndrome may be insidious over several years. The tumours may also be small and therefore difficult to locate.

In patients with small cell carcinoma of the lung, the onset of symptoms is usually rapid, making the diagnosis easier. Tumours are aggressive and the primary malignancy frequently kills the patient before significant or severe effects of glucocorticoid excess supervene. The peak incidence of ACTH syndrome is in the fourth and fifth decades and the condition is twice as common in men as in women.

Non-adrenocorticotrophic hormone-dependent Cushing syndrome

The commonest cause of the syndrome is the administration of steroids for the treatment of other conditions. This iatrogenic group will not be considered further.

Adrenal tumours

Approximately 20% of cases of Cushing syndrome are caused by either benign adrenal adenoma (10%) or frankly invasive adrenocortical carcinoma (10%). Benign adenoma is more common in adults and malignancy more common in the young. Excessive glucocorticoid secretion suppresses ACTH production from the pituitary, resulting in atrophy of the contralateral adrenal gland.

Macronodular adrenal cortical hyperplasia

This unusual condition, in which ACTH-independent cortisol secretion results from bilateral adrenal nodular hyperplasia, is sometimes seen in the young and may be associated with severe osteopaenic bone disease. The cause is unknown.

Primary pigmented nodular adrenal cortical disease

Primary pigmented nodular adrenal cortical disease producing autonomous adrenal hyperfunction occurs in conjunction with cardiac, breast and cutaneous myxomas. The adrenal nodules in this rare condition are normally small and the adrenal glands may even appear to be normal on imaging. Lentigines are found on the face, lips and conjunctivae. Pituitary adenomas and neuroectodermal tumours also occur.

Clinical features of Cushing syndrome

The most common features of Cushing syndrome are truncal obesity, weight gain, facial plethora, hirsutism, mild or moderate hypertension, menstrual irregularities, bruising and abdominal striae (Figure 20.2).

Obesity is caused by excessive cortisol stimulating appetite and promoting gluconeogenesis, which releases glucose for fat synthesis. Patients develop plethoric, rounded 'moon-like' facies, whereas truncal fat distribution results in interscapular fat pad formation. In contrast, the toxic effect of cortisol on muscles leads to muscle wasting and weakness. The proximal limb muscles are affected most and patients may develop a 'lemon on a stick' appearance.

Skin changes result from excessive cortisol depleting skin collagen. Thinning of the skin, striae and excessive bruising (either spontaneously or with minor trauma) are all features. In ACTH-dependent Cushing syndrome, high levels of ACTH melanocyte-stimulating hormone cause skin pigmentation. This is more commonly encountered in ectopic ACTH syndrome due to malignancy and Nelson syndrome (high ACTH levels following bilateral adrenalectomy for pituitary disease).

Hirsutism and acne from excessive androgen secretion are often distressing features for the patient. Adrenocortical carcinoma should be suspected, particularly if virilisim is marked.

Poor wound healing is a direct outcome of decreased collagen synthesis and increased protein catabolism. This is an important consideration in all patients in whom surgery is contemplated, particularly when an anterior abdominal approach is employed.

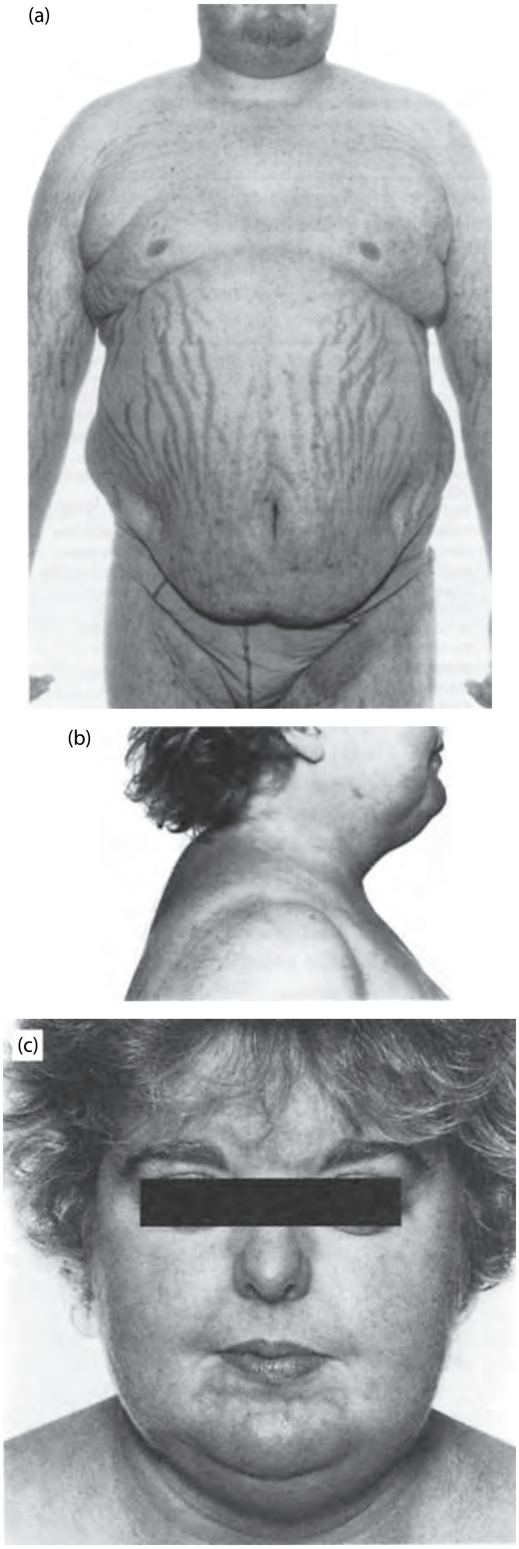


Figure 20.2 Cushing syndrome. (a) Marked truncal obesity, abdominal striae and wasted limbs. (b) Interscapular fat pad. (c) Plethoric moon face.

Hypertension and oedema are a consequence of the glucocorticoid effects of cortisol resulting in salt and water retention. Congestive cardiac failure is present in 20% of patients. Untreated this is a major cause of mortality in Cushing syndrome.

Diabetes mellitus develops in approximately 15% of patients because of the action of cortisol on carbohydrate metabolism. An abnormal glucose tolerance test, however, may be present in 70% of patients.

Osteoporosis is common (70%) and may lead to pathological fractures, particularly in the ribs, and also vertebral collapse. Urinary calcium excretion is raised (hypercalciuria), leading to renal stone formation and hypocalcaemic alkalosis.

Menstrual disturbance, male impotence, psychiatric illness and growth retardation (particularly in children) are also associated features of Cushing syndrome.

Differential diagnosis

The clinical features of Cushing syndrome may be mistakenly confused with those of the obese diabetic, particularly when there is a plethoric face or hirsutism. Polycystic ovary syndrome, with sterility and hirsutism, also poses diagnostic difficulties, but Cushing syndrome rarely causes hirsutism alone. Alcoholic patients can present with pseudo-Cushing syndrome, the signs and symptoms of which resolve following cessation of excessive alcohol intake.

Biochemical investigation of Cushing syndrome

Cushing syndrome will be initially suspected on the basis of the above clinical features. It is then necessary to confirm the presence of hypercortisolism, and subsequently identify the cause of excessive cortisol production:

- *Plasma cortisol levels* (under ACTH control) in normal individuals exhibit a diurnal variation. The highest levels are recorded at 09.00 (<140–800 nmol/L) and the lowest at midnight (<190 nmol/L). Loss of this diurnal variation is frequently exhibited in Cushing syndrome. Plasma cortisol levels recorded at 09.00 and particularly at midnight are elevated above normal limits. False-positive and false-negative observations may be caused by acute illness or depression. Even the relatively minor trauma of venepuncture can falsely elevate midnight levels.
- *Urinary free cortisol* in 24 hour urine collection is invariably raised in Cushing syndrome. Normal values are <360 nmol/day for men and <280 nmol/day for women.
- *Dexamethasone suppression tests.* In normal individuals ACTH secretion and cortisol production may be suppressed by the administration of an exogenous corticosteroid (dexamethasone). Although patients with Cushing syndrome may also exhibit ACTH suppression, it is far less marked than in normal individuals. *Overnight dexamethasone test:* 2 mg of dexamethasone (a powerful corticosteroid) is administered orally at midnight and plasma cortisol levels are measured at 09.00. When results of the overnight test are equivocal a *low-dose dexamethasone* test may be performed: plasma cortisol levels are measured prior to oral administration of dexamethasone 0.5 mg 6 hourly for 2 days. Suppression of plasma cortisol levels is seen in normal patients but rarely in patients with Cushing syndrome.
- *Insulin-induced hypoglycaemia* fails to produce a rise in plasma cortisol levels in patients with Cushing syndrome. The test is useful in differentiating between patients with the syndrome and those with depression, who may have elevated basal cortisol levels.

Identification of the cause of cortisol excess

Having confirmed a diagnosis of Cushing syndrome, the following tests are used to establish whether the disease is ACTH or non-ACTH dependent (i.e. pituitary or non-pituitary in origin):

- *Plasma ACTH measurement by radioimmunoassay.* Patients with adrenal tumours have low or undetectable levels of plasma ACTH (<10 pg/mL) compared with patients with pituitary disease, who tend to have elevated levels, although in some patients with pituitary disease ACTH levels may be in the normal range. In patients with ectopic ACTH syndrome, levels may be very high, sometimes in excess of 200 pg/mL.
- *High-dose dexamethasone test.* Patients are given 2 mg of dexamethasone 6 hourly for 2 days. In nearly all patients with pituitary disease (ACTH dependent) serum cortisol levels are suppressed. In contrast, patients with non-ACTH-dependent disease (adrenal tumour, adrenal nodular hyperplasia and ectopic ACTH syndrome) show little or no suppression. False-positive and -negative results do occur and are more likely to be encountered in patients with adrenal nodular hyperplasia and ectopic ACTH syndrome than in patients with adrenal tumours.
- *Metyrapone test.* This test is now rarely performed. Metyrapone is an 11 -hydoxylase inhibitor that blocks the final step in cortisol synthesis, reducing plasma cortisol levels, stimulating ACTH secretion and increasing production of the cortisol precursors (compound S). The test is performed by giving patients Metyrapone 750 mg (125 mg/4 hourly). A 24 hour urine collection is performed on the day before, the day of and the day after Metyrapone administration and 17-hydroxycorticosteroid (17-OHCS) levels are measured. In normal patients and patients with pituitary disease (ACTH dependent) urinary excretion of 17-OHCS is increased. In contrast, the majority of patients with adrenal tumours exhibit a decrease in 17-OHCS urinary excretion. Patients with nodular hyperplasia or ectopic ACTH syndrome usually demonstrate an impaired or absent response to Metyrapone.

Choice of test in Cushing syndrome

When Cushing syndrome is suspected clinically, diagnosis is usually confirmed by:

- measuring plasma cortisol (09.00 and midnight)
- measuring urinary free cortisol
- performing a low-dose dexamethasone test.

Confirming the cause of Cushing syndrome is best achieved by:

- measuring basal plasma ACTH
- performing a high-dose dexamethasone test
- in cases of diagnostic difficulty, performing a Metyrapone test.

Localizing procedures in Cushing syndrome

Pituitary tumour (Cushing disease)

MRI is considered the investigation of choice in localizing pituitary tumours and approximately 85% of pituitary adenomas are detectable by this technique. MRI, however, may fail to identify microadenomas of 5–10 mm in size. In these patients, bilateral selective catheterization of the petrosal sinus and ACTH measurement after CRH stimulation is the most sensitive test in confirming the presence of a microadenoma. Plain skull radiography is not helpful in identifying pituitary tumours even with thin-section tomography. CT scanning, although formally popular, will identify only approximately 50% of tumours.

Adrenal tumour

CT scanning is very effective in localizing adrenal tumours even as small as 1 cm in diameter. When an adrenal tumour is suspected as the cause of Cushing syndrome, early CT scanning may prove useful to focus other investigations. Adrenal tumours >6 cm on CT scanning are more likely to be malignant. Scanning of adrenal tumours with ^{131}I-6 -iodomethylnorcholesterol (NP59) or selenium-75 cholesterol has now been abandoned. It is unreliable and the isotopes are now unavailable in many countries (UK in particular) (Figure 20.3). MRI has increasingly become a popular modality for scanning adrenal tumours. In-/out-of-phase modality scanning may be useful in distinguishing high fat content in benign adrenal tumours (Figure 20.4).

Ectopic adrenocorticotrophic hormone

Ectopic ACTH-producing tumours are often small and may be difficult to locate. Chest radiography may identify benign or malignant bronchial tumours. CT or MRI scanning may aid in locating bronchial, pulmonary, thymic or pancreatic tumours, which may be the source of ACTH production. The adrenal glands are enlarged and usually demonstrable on CT scan (Figure 20.5).

Treatment of Cushing syndrome

Medical treatment

Patients with Cushing syndrome (either pituitary or adrenal in origin) in whom surgery is contraindicated may be treated by long-term Metyrapone therapy. This 11 -hydroxylase inhibitor reduces circulating levels of cortisol by inhibiting its synthesis from 11-deoxycortisol. Metyrapone is also useful in the preparation of patients with Cushing syndrome for definitive treatment such as adrenalectomy, trans-sphenoidal microsurgery or pituitary irradiation. Initial dosage is 250 mg 8 hourly, increasing over a week to 500 mg 6 hourly. In full dosage complete adrenal blockade is achieved, necessitating replacement therapy with 0.25 mg dexamethasone twice daily. Side effects of Metyrapone are often unpleasant and include nausea, vomiting, hirsutism and acne.

Other drugs have also been used in the medical management of Cushing syndrome, including ketoconazole, which reduces cortisol levels by inhibiting cytochrome P450 enzymes,

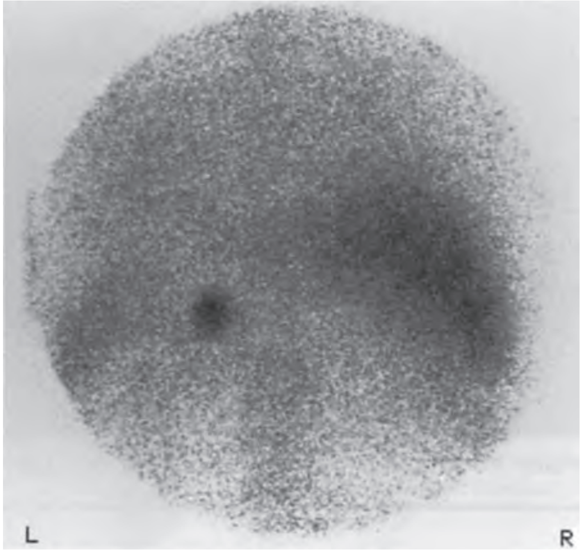


Figure 20.3 Selenocholesterol scan of a patient with Cushing syndrome due to an adenoma in the left adrenal gland. This technique has now been abandoned owing to unreliability, lack of isotopes and increased accuracy of CT/MRI.

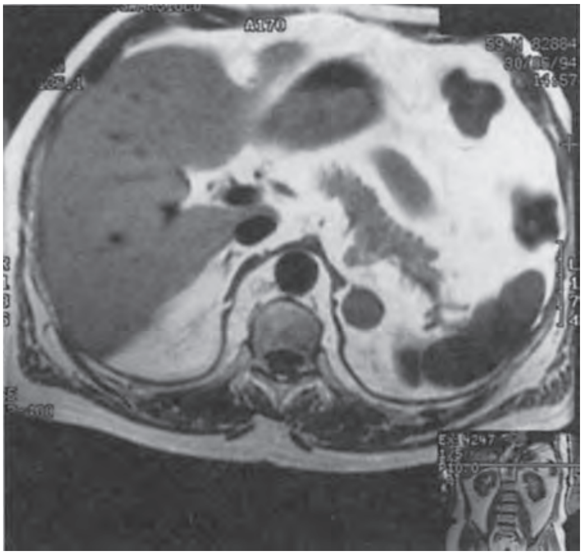


Figure 20.4 MRI scan of a patient with Cushing syndrome due to an adenoma in the left adrenal gland.

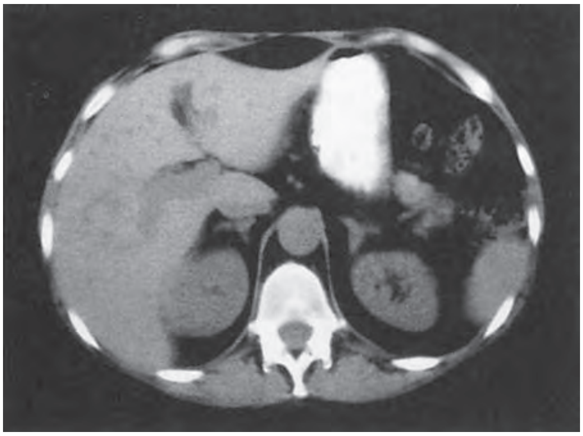


Figure 20.5 CT scan of the abdomen showing bilateral enlarged hyperplastic adrenal glands in the ectopic adrenocorticotrophic hormone syndrome.

aminoglutethimide, which inhibits the conversion of cholesterol to pregnenolone, bromocryptine, a dopamine agonist, and mitotane (discussed in Adrenocortical carcinoma).

Trans-sphenoidal microsurgery for pituitary disease

Selective removal of the pituitary basophil microadenoma by trans-sphenoidal microsurgery is currently the treatment of choice for patients with Cushing disease. When tumours have been localized preoperatively and removed completely at operation by an experienced surgeon, cure rates are high (80–90%). These patients rarely require pituitary replacement therapy. When tumours are not localized preoperatively and not readily identified at operation, partial hypophysectomy (removing two-thirds of the gland) is usually performed. Cure rates in these patients fall in comparison with the former group.

Pituitary surgery will be complicated by pituitary insufficiency when complete hypophysectomy is performed. Patients may require replacement therapy with one or more of the following agents: corticosteroid, thyroxine, testosterone, oestrogen and antidiuretic hormone (diabetes insipidus). Complications such as infection, meningitis or anosmia occur rarely.

Pituitary irradiation

Radiotherapy using either an external proton beam, linear accelerator or interstitial irradiation using gold-198 or yttrium-90 is also effective. Resolution of the clinical syndrome, however, is more protracted and may take 6–18 months. Pituitary insufficiency occurs in about 50% of patients but less frequently in children, making pituitary irradiation treatment an appropriate and acceptable therapy.

Bilateral adrenalectomy

Formerly the treatment of choice in Cushing disease, improved techniques in pituitary surgery have meant that bilateral adrenalectomy is reserved for the following situations:

- when pituitary surgery has failed or is technically not feasible
- for palliative treatment of patients with ectopic ACTH syndrome
- for patients with primary adrenal hyperplasia
- for patients with rapidly progressive or severe hypercortisolism.

Prior to surgery, treatment with Metyrapone is necessary to reduce cortisol production and correct metabolic defects. Diabetes mellitus and hypertension should be controlled if present. Appropriate venous thrombosis prophylaxis and support stockings are employed. Adequate steroid replacement therapy is essential in this procedure, the details of which are outlined earlier in this chapter. Adrenalectomized patients are dependent on lifelong glucocorticoid and mineralocorticoid replacement therapy.

A rare, late complication of bilateral adrenalectomy in patients with Cushing syndrome was first described by Nelson and colleagues in 1960. This condition is characterized by the development of an ACTH-producing pituitary tumour following bilateral adrenalectomy, accompanied by skin hyperpigmentation, headache, visual field defects and hypopituitarism. The condition may be controlled by hypophysectomy or pituitary irradiation. Prophylactic pituitary irradiation in patients undergoing bilateral adrenalectomy will reduce the incidence of Nelson syndrome.

Overall, results for bilateral adrenalectomy in the treatment of Cushing disease are good. Montgomery and Welbourn (1978) reported 65% of patients in remission at 5–15 years following surgery with 50% alive at 20 years.

Unilateral adrenalectomy

When Cushing syndrome is caused by an adrenal adenoma the treatment of choice is unilateral adrenalectomy, with most endocrine surgeons currently favouring a laparoscopic approach. The contralateral adrenal gland is suppressed in this condition and patients require steroid support both peroperatively (at tumour removal) and postoperatively. Complete weaning off steroid replacement may take more than a year. Cure rate and long-term survival are excellent.

Summary

The causes and treatment of Cushing syndrome are summarized in Figure 20.6.

Adrenocortical carcinoma

Pathology

Adrenocortical carcinoma is a rare tumour responsible for 10% of all cases of Cushing disease with a peak incidence in the fourth and fifth decades. Tumours are usually >5 cm in diameter and may be palpable. Macroscopically, they have a lobulated appearance, are greyish pink in colour and exhibit areas of haemorrhage and necrosis (Figure 20.7). They may invade local structures such as pancreas, kidney, diaphragm and bowel. Venous invasion is also common and may be gross to involve the inferior vena cava. Microscopically tumours display numerous mitoses, nuclear pleomorphism and vascular invasion. Sometimes the true malignant potential may not be apparent on conventional histological appraisal.

Adrenocortical carcinomas are aggressive malignancies with a high recurrence rate after surgery and a poor response to radiotherapy. Distant metastases to liver, lungs, bone and skin are common and are often found at the time of presentation. Most forms of treatment are disappointing and only a minority of patients survive beyond 2–3 years.

Tumours are divided into two groups: functioning (60%) and non-functioning (40%). The former account for 10% of all cases of Cushing syndrome. Functioning tumours are more common in younger patients (<40 years) and exhibit a female to male ratio of 4:1. The clinical picture is dependent upon the hormone or hormones secreted. In contrast, non-functioning tumours are more common in older patients (>40 years) and exhibit a female to male ratio of 2:1 but may produce steroid precursors such as pregnenolone.

Surgical staging

Adrenocortical carcinomas may be staged into the following groups:

- Stage I: tumour <5 cm in diameter, with no local invasion, nodal or distant metastases.

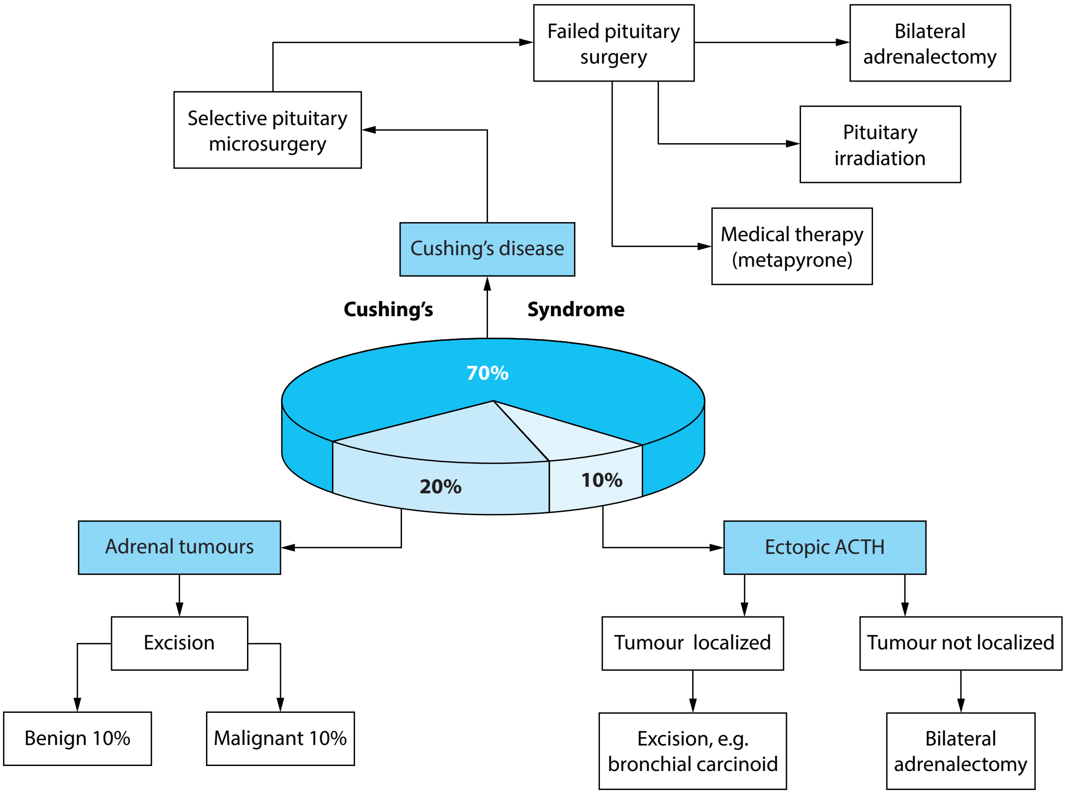


Figure 20.6 Causes and treatment of Cushing syndrome. ACTH, adrenocorticotrophic hormone.

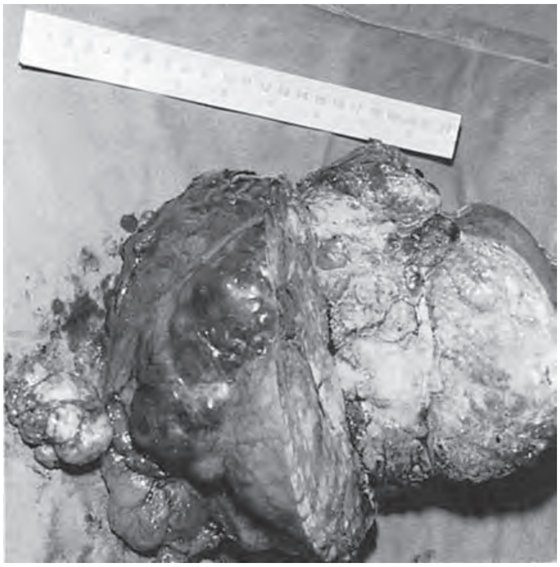


Figure 20.7 Gross specimen of adrenocortical carcinoma: haemorrhagic cut surface of a partly necrotic tumour.

- Stage II: tumour >5 cm in diameter, with no local invasion, nodal or distant metastases.
- Stage III: tumour with local invasion or local lymph node metastases but no distant metastases.
- Stage IV: tumour with local invasion *and* local lymph node *or* distant metastases.

At presentation approximately 70% of patients have either stage III or stage IV disease.

Clinical presentation

Non-functioning or minimally functioning tumours may grow to a massive size (>20 cm) before they become clinically apparent. Despite the term 'non-functioning' these tumours do produce steroid precursors and measurement of dehydroepiandrosterone may be of value in the diagnosis of malignant disease. Patients usually present with weight loss, fatigue, abdominal pain or in rare incidences with haemorrhagic necrosis of the tumour leading to acute pain, fever and/or shock. A palpable abdominal mass may be present. Functioning tumours secrete excessive steroids which may cause Cushing syndrome, hyperaldosteronism, virilization, feminization or a mixed clinical picture. Indeed a mixed hormonal secretion is highly suspicious of malignancy.

The development of Cushing syndrome in a child is likely to be due to an adrenocortical carcinoma. It occurs in association with virilization, acne and amenorrhoea in the young female, virilization in a prepubital female and feminization in the male. In contrast to benign secreting tumours the onset of symptoms is likely to be rapid. Functioning tumours may also be large at presentation. Adrenocortical cancers are not ACTH dependent and, like benign adrenocortical tumours, fail to suppress with high doses of dexamethasone.

Preoperative imaging with CT or MRI will provide valuable information regarding tumour size, local invasion and distant metastases (Figure 20.8a,b). The presence of inferior caval invasion can also be determined by MRI, rendering caval venography unnecessary (Figure 20.8c).

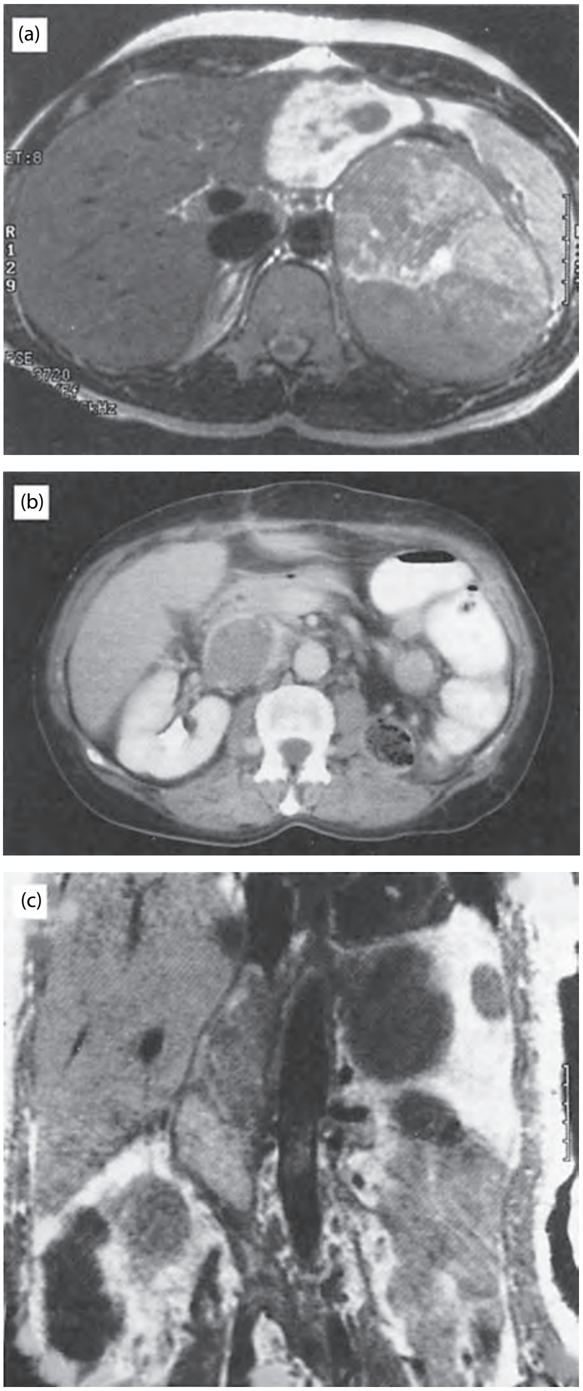


Figure 20.8 (a) MRI scan of a patient with a massive left adrenocortical carcinoma. (b) CT scan showing the tumour within the lumen of the inferior vena cava (IVC). (c) MRI scan showing tumour invasion of the IVC.

Treatment

Surgical resection offers the only prospect of cure. Unfortunately most patients at presentation have advanced incurable disease and only palliation is possible. In patients with stage I or II disease, aggressive surgical resection in the form of an *en bloc* resection, sacrificing adjacent organs such as the spleen, kidney and distal pancreas, is necessary to achieve adequate clearance. Occasionally large tumours require a thoracoabdominal approach and clearance of an involved inferior vena cava may be best achieved using cardiopulmonary bypass. Should local recurrence occur, further surgical resection has been shown to lead to prolonged survival. Steroid replacement should be given at the time of adrenalectomy for functioning tumours and maintained postoperatively.

These tumours demonstrate little radiosensitivity and radiotherapy has proved disappointing. Chemotherapy with mitotane (o,p'-DDD) has been of value particularly when used in patients prior to the development of metastatic disease following surgical resection for stage I and II disease and in an individual subsequently undergoing repeated surgical resection for recurrent disease. It is unlikely however that this agent significantly improves overall survival. Mitotane induces selective necrosis of the zona fasiculata and zona reticularis and results in destruction of the contralateral adrenal gland. Replacement steroid therapy is therefore essential. Metyrapone is useful in controlling the symptoms of cortisol excess in patients who are unfit or unsuitable for surgery.

Survival rates for adrenocortical carcinoma are poor and are dependent on disease stage at presentation. The overall 5 year survival is as low as 16% but stage I and II disease may have a survival of more than 50% compared with 0% for those with stage IV tumours.

Aldosteronism

Primary aldosteronism

Primary hyperaldosteronism is a rare syndrome characterized by hypertension, hypokalaemia, hypernatraemia and suppressed plasma renin activity (PRA). It is caused by excessive aldosterone secretion from the zona glomerulosa of the adrenal cortex. This potentially curable condition accounts for <1% of all patients with hypertension and was first described by Dr Jerome Conn in 1955, 3 years after the hormone aldosterone had been identified.

Excessive aldosterone production causes expansion of plasma volume and elevation of the blood pressure (see Physiology). The rise in blood volume and sodium ion concentration is detected by the juxtaglomerular apparatus and renin secretion falls in response. Loss of potassium and hydrogen ions in the urine leads to hypokalaemia and metabolic alkalosis.

Adrenocortical adenoma (Conn's tumour) is by far the most common cause of primary hyperaldosteronism, accounting for approximately 85% of cases. Other less common causes are bilateral adrenocortical hyperplasia (idiopathic hyperaldosteronism), aldosterone-producing adrenocortical carcinoma, glucocorticoid-suppressible hyperaldosteronism (familial type 1), non-glucocorticoid-suppressible hyperaldosteronism (familial type 2, which may be an adenoma or hyperplasia) and aldosterone-producing ovarian carcinoma.

Pathology

Conn's tumours are classically unilateral, <2 cm in size, project from the adrenal gland and have a characteristic golden or canary yellow appearance (Figure 20.9). Tumours are composed of lipid-laden clear cells and occur more frequently on the left side. Bilateral tumours are found in 10% of patients.

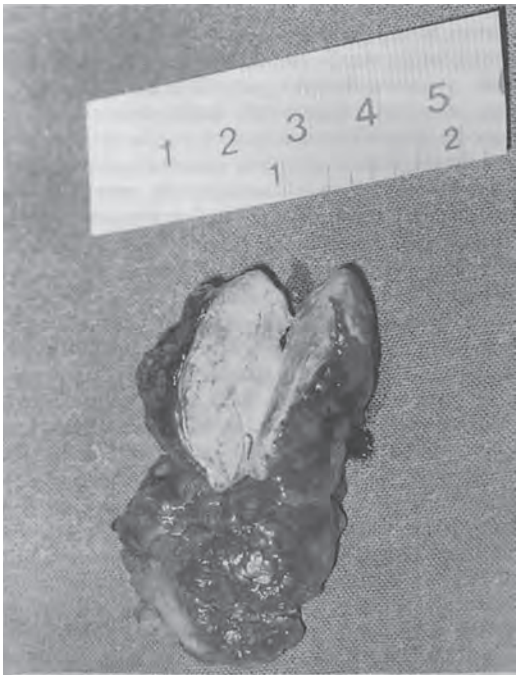


Figure 20.9 Conn's adrenal adenoma – gross specimen.

The remainder of the adrenal cortex in patients with Conn's tumours is not always normal. Microscopic and/or macroscopic nodular hyperplasia is a frequently associated feature, present in approximately 40–50% of patients.

Idiopathic hyperaldosteronism is found in 10–15% of patients with primary hyperaldosteronism. The disease is bilateral and composed of both diffuse and focal areas of microscopic and macroscopic hyperplasia. Nodule formation is frequently present, probably caused by hypertensive vascular changes. The nodules are capable of producing cortisol but not aldosterone. The dividing line between solitary adenoma, adenoma in a background of zona glomerulosa hyperplasia and hyperplasia with no dominant nodule is somewhat blurred. Indeed variations may represent a spectrum of disease. Identifying the cause of primary hyperaldosteronism is crucial and has a direct bearing on appropriate management.

Clinical presentation

The syndrome of primary hyperaldosteronism is characterized by continuous hypertension, which is often severe but rarely malignant. Patients may present at any age but more commonly between the third and fifth decades. Conn's tumours occur twice as commonly in women, although idiopathic hyperplasia is equally distributed between sexes and tends to occur at a later age.

The duration of hypertension prior to diagnosis is reported to be around 7 years and may be resistant to usual antihypertensive medication. When hypokalaemia is marked, patients may experience muscle weakness, cramps, headaches, polydipsia, polyuria and nocturia. In rare cases where hypokalaemia is particularly severe, patients may experience episodes of intermittent flaccid paralysis or even tetany. Symptoms of hypokalaemia may be brought on by the administration antihypertensive diuretics, particularly the thiazides. On rare occasions excessive salt intake may also cause episodes of hypokalaemia. Patients may be symptomless and the diagnosis is often only suspected when routine biochemical analysis characteristically reveals hypokalaemia associated with mild hypernatraemia.

Biochemical investigation and diagnosis

When hyperaldosteronism is suspected clinically, confirmation of the diagnosis is obtained by demonstrating elevated plasma aldosterone concentration (PAC) with suppressed PRA. This contrasts with secondary hyperaldosteronism, where both PAC and PRA are raised. Prior to carrying out biochemical studies potassium stores should be replenished (hypokalaemia inhibits aldosterone secretion) and drugs affecting renin–aldosterone regulation should be discontinued for 4–6 weeks. This includes spironolactone, ACE inhibitors and all diuretics. In patients in whom hypertension is marked, antihypertensive medication may be continued with agents such as prazosin and guanethidine, although calcium channel blockers and beta-blockers probably do not significantly affect results.

Primary hyperaldosteronism is confirmed if in a hypertensive patient PAC is increased (normal 2.2–15 ng/dL) in association with PRA being suppressed below normal (0.2–0.5 ng/dL). It has been suggested that a PAC/PRA ratio of >50 is diagnostic. Urinary potassium excretion should be in excess of 40 mmol/day.

Aldosterone suppression test

When results are equivocal further evaluation is deemed necessary. Diagnosis of primary hyperaldosteronism in this situation may be confirmed by failure to demonstrate suppression of urinary aldosterone secretion in response to a sodium load. Caution is necessary in performing these tests as biochemical disturbances can be severe and marked hypokalaemia may ensue. Patients should therefore be normokalaemic prior to testing and have potassium supplementation throughout the test. Oral sodium loading takes place over 3 days at a dose of 9 g/day; on the third day a 24 hour urine collection is made and urinary aldosterone, potassium and sodium levels are measured. The 24 hour urinary sodium excretion should exceed 200 mEq (documenting adequate sodium loading) and the diagnosis of hyperaldosteronism is confirmed if aldosterone levels exceed 12   g.

Differentiation between adenoma and idiopathic hyperplasia

Following diagnosis of primary hyperaldosteronism it is necessary to distinguish between an aldosterone-producing adenoma (APA) and idiopathic hyperplasia (IHA). This distinction has important therapeutic implications, as only cortical adenomas are likely to benefit from surgery. Aldosterone-producing cortical adenomas remain relatively unresponsive to angiotensin but still respond to ACTH and follow the corticotrophin circadian rhythm. In idiopathic hyperplasia there is a hypersensitivity to angiotensin II. These mechanisms can be used as the basis for a test to distinguish between the two conditions.

PAC is measured after overnight recumbency and after the patient has been ambulatory for 4 hours. In those with adenomas, PAC falls, whereas in patients with idiopathic hyperplasia PAC rises. Although this test is not absolutely reliable Young reported an overall accuracy of 85% in 246 patients in a collective review of the literature.

Reports of subtypes of primary hyperaldosteronism termed primary adrenal hyperplasia and aldosterone-producing renin-responsive adenoma (both responding to unilateral adrenalectomy) further confuse the issue.

Localization

CT scanning has proven to be reliable in localizing Conn's tumours >1 cm in size in more than 90% of patients. Identification of an obvious adrenal cortical adenoma increases confidence of the diagnosis of Conn's tumour versus idiopathic hyperplasia in patients with primary hyperaldosteronism. Tumours are usually homogeneous with decreased attenuation both before and after contrast enhancement because of their low lipid content. Most tumours measure <2 cm but those smaller than 1 cm may be missed, particularly in thin patients with little perinephric fat. MRI has also proved useful in identifying tumours and may well become the method of choice (Figures 20.10 and 20.11).

The most accurate functional method of localizing Conn's tumours remains selective venous catheterization (Table 20.1). The technique, however, is invasive and potentially hazardous, with reported extravasation of contrast, intra-adrenal haemorrhage and adrenal necrosis; these complications however are not common.

As the functioning adenoma may be small (4–8 mm) and incidental non-functioning adrenal adenomas, relatively common. Localization with selective venous catheterization is advocated in all patients with Conn's, to avoid the mistake of removing non-functioning adenomas on the wrong side. Difficulty in placing the catheter into the right adrenal vein (which is short and at 90° to the cava) may lead to sampling errors. These errors can be minimized by simultaneously measuring cortisol and employing the aldosterone–cortisol ratio as an indicator of any true increase in hormone concentration.

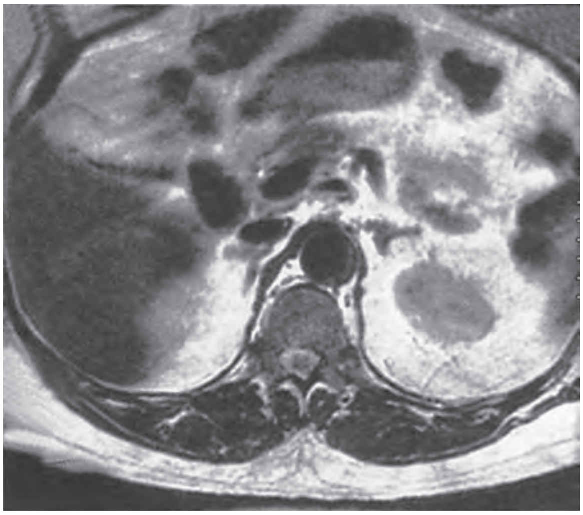


Figure 20.10 MRI scan of the abdomen in a patient with Conn syndrome due to a microadenoma in the left adrenal gland.

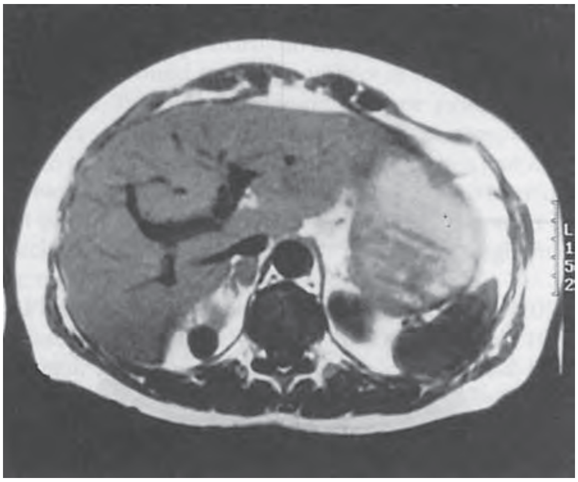


Figure 20.11 MRI of the abdomen in a patient with Conn syndrome due to right adrenal cortical adenoma.

Table 20.1 Selective venous sampling results in the patient with microadenoma shown in Figure 20.10

Vein	Aldosterone (A) (pmol/L)	Cortisol (C) (nmol/L)	A/C ratio
Right adrenal	2910	5300	0.5
Left adrenal	17 000	1810	9.4
Peripheral	660	320	2.1

Treatment

Once an aldosterone-secreting cortical adenoma is confirmed biochemically and localized by CT or radioisotope scan the treatment of choice in patients deemed medically fit for surgery is unilateral adrenalectomy. Traditionally an open posterior approach has been used, but this is a tumour which lends itself to laparoscopic adrenalectomy very well and is now currently the preferred technique for most endocrine surgeons. Patients are prepared for surgery by correction of hypokalaemia and hypertension, usually with spironolactone 100 mg/day. Those in whom hypertension responds well to spironolactone preoperatively tend to have a more favourable outcome from surgery.

Patients with idiopathic hyperplasia, those in whom hypertension persists following adenoma excision and those who are unfit for surgery because of medical reasons should be treated with spironolactone. Potassium-sparing diuretics also have a significant effect in controlling blood pressure and restoring potassium balance.

Postoperative follow-up

Unilateral adrenalectomy for adenomas has been reported to improve hypertension in the majority and to achieve normotension in 44–98% of patients, although the average cure rate appears to be about 70% at 1 year. Some patients may have persistent hypertension immediately postoperatively and others develop further hypertension over a period of time.

Sex and age have been demonstrated to be the only two significant prognostic factors in determining operative success. The older the patient, the less likely adrenalectomy is to restore normotension. This is probably because long-term hypertension causes irreversible pathological changes in blood vessel walls and coexisting causes of hypertension are also present. Hypertension is also more likely to be labelled 'essential' in elderly patients than in younger ones, which may lead further to a delay in diagnosis and thus long-term vessel damage. Women also have a more favourable response to adrenalectomy than men, possibly because female hormones confer a protective effect on blood vessels.

Summary

The investigation and treatment of primary aldosteronism is summarized in Figure 20.12.

Secondary hyperaldosteronism

Severe hypertension from cardiovascular or renal disease may be associated with excessive secretion of aldosterone termed secondary hyperaldosteronism. Hypertensive stimulation of the juxtaglomerular apparatus produces renin, which in turn leads to aldosterone secretion via the renin–angiotensin mechanism. Raised PRA therefore distinguishes secondary hyperaldosteronism from primary hyperaldosteronism (in which PRA is suppressed), this being an important clinical distinction.

Phaeochromocytoma

Phaeochromocytoma is a functioning tumour of the catecholamine-producing chromaffin cells. Derived from the Greek *phaeo* meaning dusky and *chroma* meaning colour, the tumour is characterized by a syndrome resulting from excess catecholamine production. The condition was first described in 1886 by Frankel, following a postmortem that had demonstrated bilateral adrenal tumours in an 18 year old patient who had presented with hypertension, sudden collapse and death. The term phaeochromocytoma was first used by Pick in 1912 but it was not until 1926 that the first successful

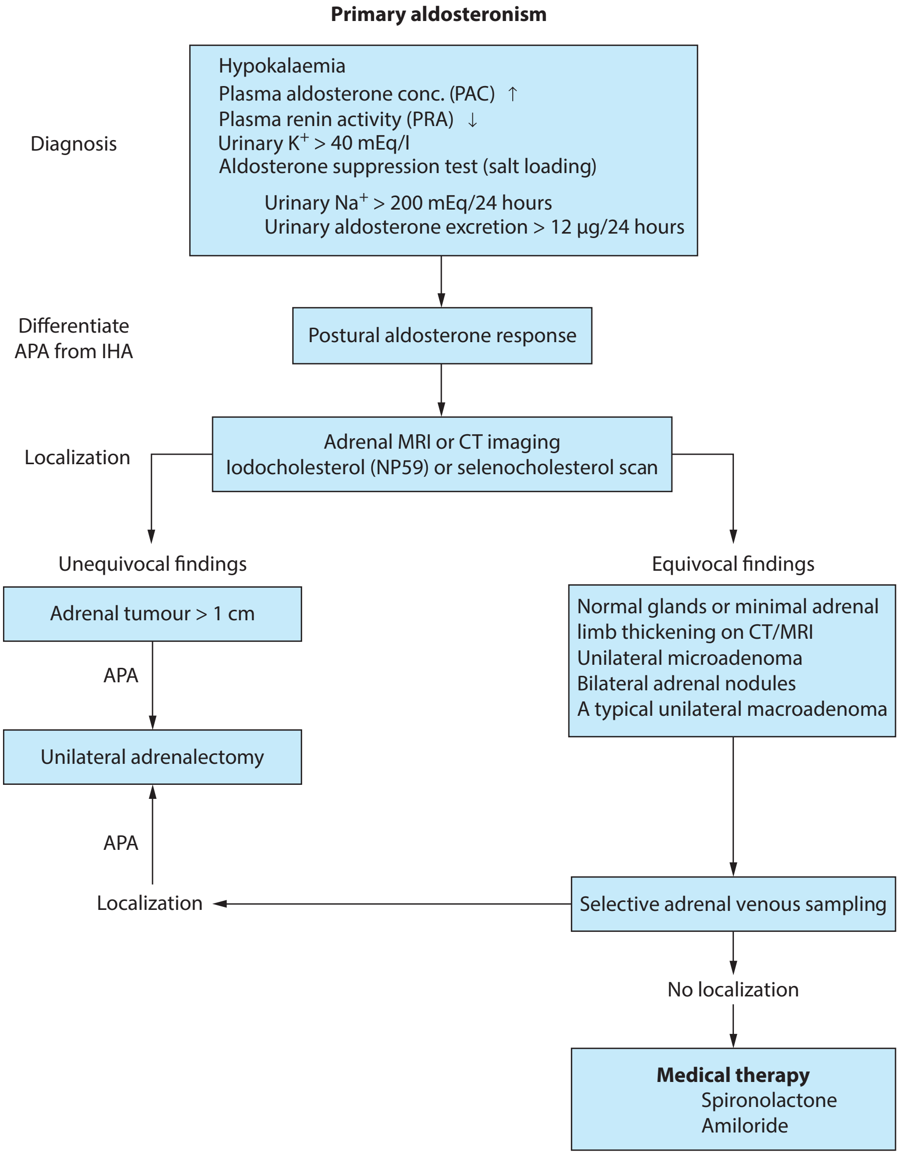

Primary aldosteronism
Diagnosis
Hypokalaemia
Plasma aldosterone conc. (PAC) ↑
Plasma renin activity (PRA) ↓
Urinary K+ > 40 mEq/l
Aldosterone suppression test (salt loading)
Urinary Na+ > 200 mEq/24 hours
Urinary aldosterone excretion > 12 μg/24 hours
Differentiate APA from IHA
Postural aldosterone response
Localization
Adrenal MRI or CT imaging
Iodocholesterol (NP59) or selenocholesterol scan
Unequivocal findings
Adrenal tumour > 1 cm
APA
Unilateral adrenalectomy
APA
Localization
Equivocal findings
Normal glands or minimal adrenal limb thickening on CT/MRI
Unilateral microadenoma
Bilateral adrenal nodules
A typical unilateral macroadenoma
Selective adrenal venous sampling
No localization
Medical therapy
Spironolactone
Amiloride

Figure 20.12 Investigation and treatment of primary aldosteronism. APA, aldosterone-producing adenoma; IHA, idiopathic hyperaldosteronism.

surgical removal of such a tumour was performed by Roux in Lausanne, a feat repeated by C.H. Mayo in the USA during the following year.

In recent years, advances in biochemical diagnosis, accurate preoperative localization, preparation with -and -adrenergic receptor-blocking agents and enhanced surgical skills have meant that patients with phaeochromocytoma can expect a favourable outcome from appropriate treatment.

Pathology

Most tumours are unilateral, more common on the right side and range considerably in size from 1 cm to more than 15 cm. Weight varies accordingly but usually tumours are between 50 and 200 g. Sporadic tumours are usually unilateral and unifocal. In contrast, familial tumours are more likely to be bilateral (50–70%) and multifocal.

Macroscopically, tumours are well encapsulated, frequently cystic-containing areas of haemorrhage and necrosis. When sectioned the cut surface is pinkish grey in colour and highly vascular.

Chromaffin cells are of neuroectodermal origin and although located primarily in the adrenal medulla they may also be found in small numbers in the sympathetic ganglia. Tumours developing in these extra-adrenal sites are termed *catecholamine-secreting paragangliomas.* These tumours are more common in children and may be located along the course of the sympathetic chain but predominate at the aortic bifurcation (the organ of Zuckerkandl). Approximately 40% of paragangliomas are malignant.

Phaeochromocytoma has been termed the ‘10%’ tumour. Approximately 10% are extra-adrenal, 10% bilateral, 10% malignant, 10% are in children and 10% are familial, associated with other genetically determined conditions such as multiple endocrine neoplasia (MEN) types 2A and 2B, von Hippel–Lindau disease or von Recklinghausen’s neurofibromatosis.

Histologically, diagnosis of malignancy in these tumours is rarely possible as many lesions, both benign and malignant, exhibit evidence of capsular and vascular invasion. The only reliable indicator of malignancy is demonstration of local invasion into adjacent structures such as the renal or inferior vena caval veins, or localization of tumour in sites other than the sympathetic chain such as bone, lung and liver.

Incidence

Phaeochromocytoma may be found at postmortem in approximately 0.1% of hypertensive patients. The incidence in the population is of the order of one or two cases per 100 000, and approximately 800 deaths per year occur in the USA from phaeochromocytoma. The condition remains undiagnosed in 35% of sufferers during their lifetime. In patients in whom phaeochromocytoma is not diagnosed, death may ensue from myocardial infarction or cerebrovascular accident during or immediately after even a minor surgical operation. A high index of suspicion is therefore necessary to make the diagnosis and avoid this potential catastrophe.

Clinical presentation

Hypertension is by far the most common manifestation, mimicking ‘essential’ hypertension by being continuous in 50% of patients but paroxysmal in the remaining 50%. Excess norepinephrine (noradrenaline), epinephrine (adrenaline) or dopamine secretion predisposes to hypertension usually associated with palpitations, sweating and headache. So commonly are these symptoms associated with phaeochromocytoma that the absence of all three has been said to rule out the diagnosis in patients with hypertension.

Events which raise intra-abdominal pressure such as bladder emptying, sexual activity, defaecation or exercise may trigger attacks that may last from several minutes to several hours. Other less common causative factors are alcohol, drugs (notably tricyclic antidepressants) and foods containing high levels of tyramine. In undiagnosed patients, labour or invasive procedures such as angiography may precipitate particularly severe hypertensive episodes resulting in peripheral circulatory failure and even cardiac arrest.

The incidences of myocardial infarction, hypertensive encephalopathy, cerebrovascular accident and catecholamine cardiomyopathy are all raised in these patients. Other manifestations of phaeochromocytoma, resulting from increased levels of catecholamines, are flushing, pallor, pupillary dilation, Raynaud’s phenomenon, fever, tremors, nausea, weakness, vertigo, anxiety and an impending sense of doom. Gastrointestinal manifestations may be generalized abdominal pain, pseudo-obstruction, ileus, ischaemic enterocolitis and acute megacolon. Malignant tumours may also be responsible for glucose intolerance and severe hyperglycaemia.

Physical examination frequently reveals a thin, pale, anxious patient. Hypertension may be severe and, in contrast to essential hypertension, 70% of patients exhibit a postural drop in blood pressure usually associated with concomitant tachycardia. An abdominal mass may be present in up to 15% of patients, palpation of which may precipitate paroxysmal symptoms and should be conducted with care in patients in whom there is a high index of suspicion. Clinical features such as a thyroid nodule, Marfinoid habitus or neurofibromatosis may be suggestive of familial disease such as MEN 2A, MEN 2B or von Recklinghausen’s disease.

The differential diagnosis includes ‘essential’ and ‘labile’ hypertension, anxiety neurosis and other psychiatric disorders, hyperthyroidism, diabetes mellitus and functional bowel disorder. It is not surprising therefore that phaeochromocytoma has been labelled the great ‘mimic’ with all the attendant diagnostic difficulties.

Investigation and diagnosis

Diagnosis of phaeochromocytoma is provided by establishing catecholamine excess, followed by localization of the tumour. Armstrong and colleagues described the metabolic pathways of catecholamine production and excretion, which provide the basis for biochemical investigation of phaeochromocytoma. Dopamine, norepinephrine and epinephrine are formed in the chromaffin cells of the adrenal medulla from phenylalanine and tyrosine. Norepinephrine and epinephrine are converted to normetadrenaline and metadrenaline prior to excretion in the urine. These metanephrines are then further converted to vanillyl mandelic acid (VMA), also excreted in the urine.

Diagnosis is thus established by 24 hour urine collection in an acidified container and measurement of the metanephrine and VMA content of the urine. Epinephrine, norepinephrine and dopamine are also assayed by high-pressure liquid chromatography, reducing the possibility of drug interactions producing spurious results. In borderline cases, measurement of plasma free catecholamine levels may be helpful during acute hypertensive attacks but is otherwise unnecessary. Pharmacological provocation tests are potentially dangerous and are only of historic interest.

Localization

Once the diagnosis of phaeochromocytoma has been confirmed the exact site of the tumour or tumours must be identified. MRI is now established as the localizing procedure of choice, in preference to CT (Figures 20.13–20.15). Tumours are displayed as low-signal intensity T_1 weighted images with a characteristic, almost unique, hyperintense signal on T_2 weighted images. This contrasts with non-functioning adrenal adenomas, which have the same signal characteristics as normal adrenal tissue. MRI is of particular value in locating extra-adrenal paragangliomas, especially those located in the urinary bladder and paracardiac regions (Figure 20.16). MRI negates the need for contrast enhancement of tumours necessary with CT imaging, thus avoiding the potential hypertensive crises which contrast material can precipitate, although a recent report suggest that use of non-ionic contrast material is probably safe.

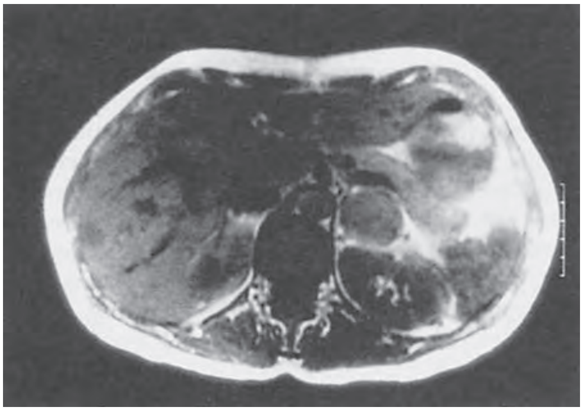


Figure 20.13 MRI scan of left adrenal phaeochromocytoma.

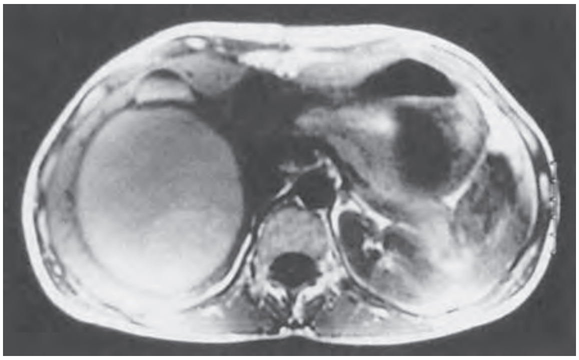


Figure 20.14 MRI scan (T_2 weighted) of a large right adrenal phaeochromocytoma. This lesion was originally referred as a possible hepatic cyst.

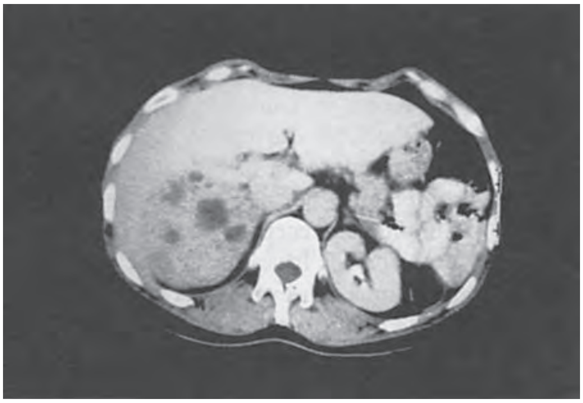


Figure 20.15 CT scan of malignant phaeochromocytoma in the right adrenal gland. Typical heterogeneous appearance with areas of haemorrhage.

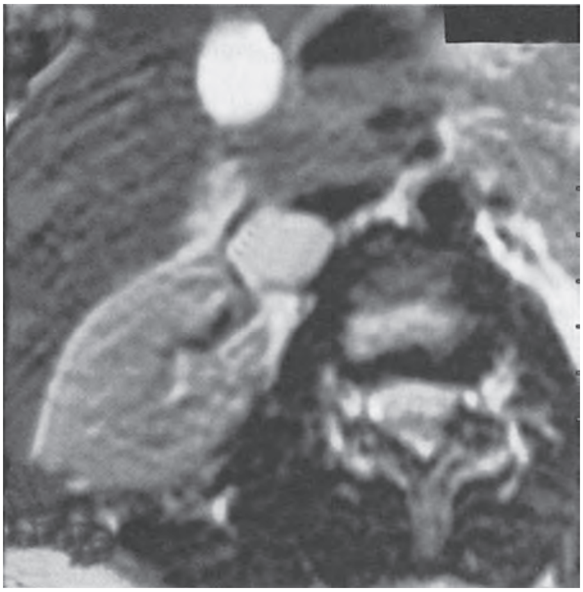


Figure 20.16 MRI scan (T_2 weighted) of an extra-adrenal phaeochromocytoma situated close to the right renal hilum.

When MRI or CT has failed to accurately localize the tumour ^{23}I-metaiodobenzylguanidine (MIBG) or indium-111 pentetreotide may be useful (Figures 20.17 and 20.18). MIBG has been reported to have a sensitivity of 90% and a specificity of 100% in tumour localization, although to achieve this it may be necessary to repeat the scan several times in the 72 hour period after administration of the tracer. Invasive procedures such as selective venous catheterization and selective adrenal angiography are no longer deemed necessary in tumour localization.

Treatment

Catecholamine-secreting tumours are treatable causes of hypertension, and surgical resection of the tumour following

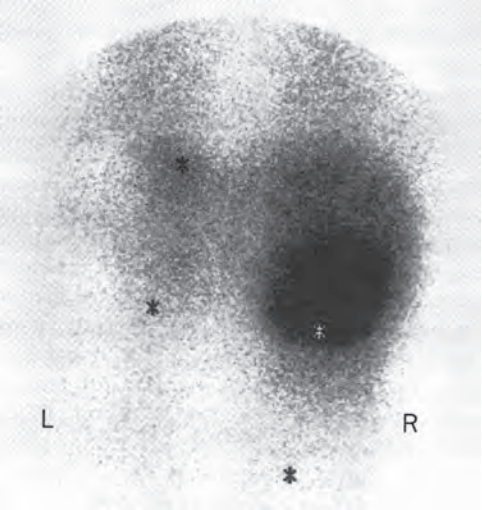


Figure 20.17 MIBG (metaiodobenzylguanidine) scan of a malignant right phaeochromocytoma.

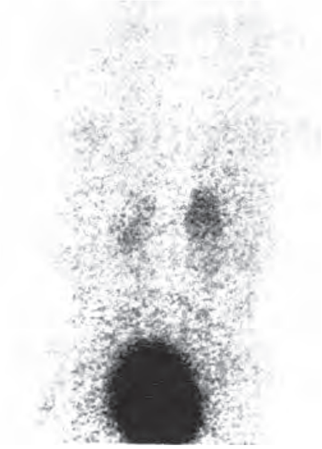


Figure 20.18 MIBG (metaiodobenzylguanidine) scan of a 5-year-old boy with bilateral phaeochromocytoma (right tumour within the adrenal gland, left tumour extra-adrenal).

pharmaceutical control of hypertension is the recommended management. Patients who have medical contraindications to surgery or patients have widespread metastatic disease from malignant phaeochromocytoma (in whom there is little hope of obtaining a surgical cure) will not be considered candidates for surgery. These patients in this group may benefit from treatment with -methyl- -tyrosine, an inhibitor of catecholamine synthesis which can reduce circulating levels by as much as 50%, offering some symptomatic control. Recent studies with the tyrosine kinase inhibitor sunitinib indicate that patients with metastatic phaeochromocytoma due to specific conditions (e.g. von Hippel–Lindau disease) may benefit from this drug.

Preoperative control of hypertension

Anaesthetic induction or even minimal manipulation of adrenal or extra-adrenal phaeochromocytomas may cause dramatic and dangerous fluctuations in blood pressure. It is therefore vital to control blood pressure preoperatively. The agent of choice in the routine patient is the long-acting -adrenergic blocker phenoxybenzamine. Medical control of hypertension in a patient thought to have a phaeochromocytoma should be commenced once the tumour is suspected, prior to any localizing procedures. Untreated patients are at significantly increased risk of myocardial infarction, cerebrovascular accident, acute heart failure, arrhythmias and sudden death.

Treatment with non-competitive alpha-blockade allows expansion of the intravascular volume, with concomitant fall in blood pressure, leading to the essential benefit of reduced frequency and intensity of peroperative hypertensive episodes. A disadvantage of alpha-blockade is the risk of hypotension following removal of the tumour and withdrawal of the -agonist activity of the catecholamines.

Phenoxybenzamine should be commenced at least 7–10 days prior to surgery. The initial dose should be 10 mg b.d., increasing by 10–20 mg/day, until hypertension is controlled and the patient experiences orthostatic hypotension. Often a dose up to 160 mg/day is required to achieve this effect. Patients should be encouraged to take liberal fluids and salt to help replace the expanded intravascular volume. Side effects of phenoxybenzamine include sedation, weakness, lassitude, nausea, nasal congestion and pedal oedema.

Three to four days prior to surgery, gentle beta-blockade with propranolol 40 mg/day is instituted. The use of preoperative beta-blockade had been somewhat controversial, although it is of particular use in patients with pure epinephrine-secreting tumours and those experiencing tachycardia and arrhythmias. Asthma and cardiac failure are absolute contraindications to the use of beta-blockers and -adrenergic blockade should not be used without prior alpha-blockade, as unopposed vasoconstriction can lead to potentially catastrophic hypertension.

The use of selective alpha$_1$-blockers, such as prazosin hydrochloride, and the newer alpha$_1$-blockers, such as terazosin and doxazosin (which selectively block postsynaptic alpha$_1$ receptors and have a shorter duration of action), may allow a reduction in blood pressure without the reflex tachycardia associated with non-competitive alpha-blockade. Labetalol, which has - and -adrenergic blocking properties, has a much shorter action than phenoxybenzamine and may be employed by anaesthetists peroperatively in controlling blood pressure. Other drugs reported to control blood pressure in phaeochromocytoma are calcium channel blockers such as nifedipine. Use is somewhat controversial but a claimed advantage of nifedipine is that patients are unlikely to experience unpleasant orthostatic hypotension.

Anaesthesia for phaeochromocytoma

Close co-operation among endocrinologist, anaesthetist and surgeon should ensure that the patient arrives in theatre in the optimal condition for surgery. Subsequently a co-ordinated team approach between surgeon and anaesthetist should ensure a safe outcome.

Patients are prepared for surgery by insertion of central venous and arterial lines. Electrocardiograph monitoring and intravenous access are essential and urinary catheterization is helpful. Swan–Ganz catheterization to monitor pulmonary wedge pressure has been advocated but its use is probably best restricted to patients in whom cardiac function is known to be seriously compromised, especially those with catecholamine cardiomyopathy.

Anaesthesia is best maintained by isoflurane rather than halothane, which has the potential for unwanted cardiac arrhythmias. A range of pressure-regulating agents (phentolamine, sodium nitroprusside, norepinephrine and dopamine) should be available in order to maintain and control blood pressure on a minute-to-minute basis throughout the procedure. Intravenous lidocaine (50–100 mg) is of value in controlling arrhythmias. Propranolol (1 mg) may also be of benefit. Cardioversion equipment must be readily available.

Danger periods during surgery for phaeochromocytoma include induction, intubation and peroperative handling of the tumour. Another crucial period may be immediately after tumour resection when hypotension can ensue. This situation is best managed by a co-ordinated effort by the surgeon and anaesthetist. Volume expansion with blood or colloids is usually sufficient but in refractory cases it may be necessary to give an epinephrine or a dopamine infusion. Failure of the blood pressure to fall following tumour removal may suggest a second tumour or undiagnosed metastatic disease. When bilateral adrenalectomy is performed the procedure must be covered with hydrocortisone 100 mg preoperatively on induction and further hydrocortisone at the time of gland removal. Intravenous hydrocortisone (100 mg 6 hourly) is continued postoperatively, reducing as full oral steroid replacement therapy is instituted.

Surgery for phaeochromocytoma

Surgical approaches for phaeochromocytoma have changed over recent years. Formerly an anterior transabdominal approach was advocated, through either a midline or a bilateral subcostal incision. This approach enabled a laparotomy to be performed to accurately identify the site of disease and to ensure that no lesion was overlooked. Transabdominal access is especially necessary in large tumours, particularly on the right side where the disease may extend posterior to the vena cava and lead to problems of vascular control. In rare instances a thoracoabdominal procedure may be necessary.

As localizing techniques improved a unilateral focused approach was advocated, usually via the posterior or posterolateral route. The advent of laparoscopic adrenalectomy and the increasing refinement of localizing procedures such as MRI has meant that now a unilateral, focused, endoscopic approach has become virtually the routine surgical method for removal of phaeochromocytomas. This may be accomplished for tumours up to 6 cm in diameter via a transabdominal laparoscopic approach (favoured by most surgeons) or a posterior route (technically more difficult), appropriate for smaller tumours.

Technical aspects are covered in the section on adrenal surgery in this chapter (Adrenalectomy) but special considerations in surgery for phaeochromocytoma include minimal handling of the tumour (made easier by a laparoscopic approach), complete haemostasis (venous ooze is a potential hazard in the alpha-blocked patient) and close communication with the anaesthetist.

Extra-adrenal phaeochromocytomas are managed by a variety of surgical approaches depending on their location. In the case of bladder tumours, partial, segmental or even total cystectomy may be necessary, particularly when tumours are suspected to be malignant.

Postoperative management

Postoperatively intensive care monitoring of arterial blood pressure, central venous pressure and urinary output is vital. Requirements for intravenous fluid replacement of blood and plasma expanders will be largely determined by the above measurements. Often large amounts of fluid will be necessary because of the relative disproportion between vascular capacitance and circulating blood volume following the removal of the chronic catecholamine vasoconstrictor drive. Residual phenoxybenzamine can also cause fluid loss into the retroperitoneal space.

When hypotension persists, haemorrhage rather than refractory vasodilation should be considered. Care should be taken to avoid overtransfusion leading to pulmonary oedema. Blood glucose levels should also be closely monitored, particularly in the first 6 hours postoperatively as dangerous, even life-threatening, hypoglycaemia may ensue, requiring glucose infusion. When hypertension persists this may be due to residual tumour, metastatic disease or chronic renal damage secondary to longstanding severe hypertension.

All patients should have urinary catecholamine and VMA measurements repeated after resection and annually thereafter. In patients with malignant disease the true nature of the condition may not become apparent for many years, thus long-term follow-up is advised and even benign tumours may occasionally recur. Overall, however, prognosis in well-prepared patients undergoing careful elective surgery is excellent, with low morbidity and mortality.

Phaeochromocytoma in special circumstances

Unsuspected phaeochromocytoma encountered during surgery

This potentially fatal condition presents a considerable challenge to both surgeon and anaesthetist. Warning signs include unexplained tachycardia, arrhythmias or hypertension during anaesthetic induction, at attempted tumour removal or for surgery for a completely unrelated condition in an unprepared patient. Should these warning signs go unheeded patients may develop cardiac failure, pulmonary oedema and fatal cardiac arrest. In this situation alpha-blockade should be instituted immediately and the procedure terminated. The patient should then undergo full investigation, preparation and localization prior to excision of the tumour at a later date.

Malignant phaeochromocytoma

Malignancy occurs in about 10% of phaeochromocytomas but this rises to 40–50% in extra-adrenal tumours. The difficulties of

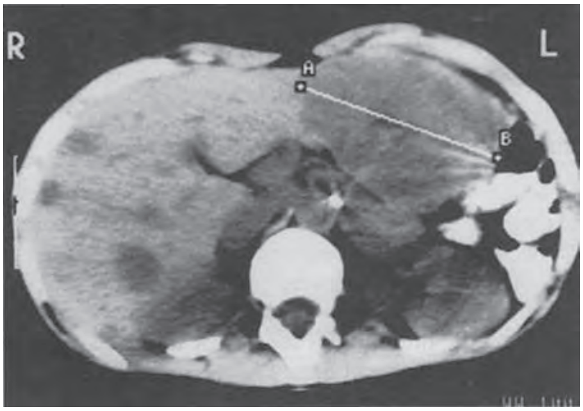


Figure 20.19 CT scan of recurrent malignant phaeochromocytoma in the left adrenal tumour bed. Multiple metastases are seen in the right lobe of the liver.

diagnosing malignancy histologically have already been referred to. Surgical resection remains the principal therapeutic measure.

Localization of metastatic disease may be aided by an MIBG scan, although this has a false-negative rate of approximately 10%. In some instances ^{131}I-MIBG may have a therapeutic as well as a diagnostic role. When possible, recurrent disease should be resected, debulked or ablated (Figure 20.19). Chemotherapy has not proved very effective but radiotherapy has been reported to occasionally provide useful palliation, particularly in bony metastases. New agents such as sunitinib (see above) may in future offer some symptomatic relief in very specific patient groups with metastatic or non-resectable disease.

Survival rates in malignant phaeochromocytoma are between 35% and 40% at 5 years, although on rare occasions patients with distant metastases have been reported to survive longer. Extra-adrenal tumours, however, carry a less favourable prognosis.

Phaeochromocytoma in pregnancy

Phaeochromocytoma in pregnancy is a potentially lethal condition: maternal mortality of 40% and fetal mortality of 40–56% have been reported. Hypertension is common in pregnancy and phaeochromocytoma may be mistaken for pre-eclampsia. Therefore, catecholamine excess should be excluded in all hypertensive pregnant women. A high index of suspicion should be maintained in pregnant women who present with unexplained cardiovascular collapse, in those who exhibit severe or labile hypertension in early pregnancy, or in patients who have a positive family history. Maternal and fetal mortality rates are greatly reduced if the diagnosis is made prior to the onset of labour as hypertensive paroxysms will be precipitated by uterine contractions, anaesthesia or caesarean section.

Following diagnosis, tumours may be localized with MRI, avoiding excessive use of radiographs.

After instigating the usual alpha-blockade and beta-blockade 2–3 days preoperatively in the first and second trimesters surgical excision of tumours may be performed. Some patients may elect for termination of the pregnancy. In the third trimester caesarean section followed by removal of the phaeochromocytoma under the same anaesthetic is appropriate (Figure 20.20). Vaginal delivery should be avoided at all costs.

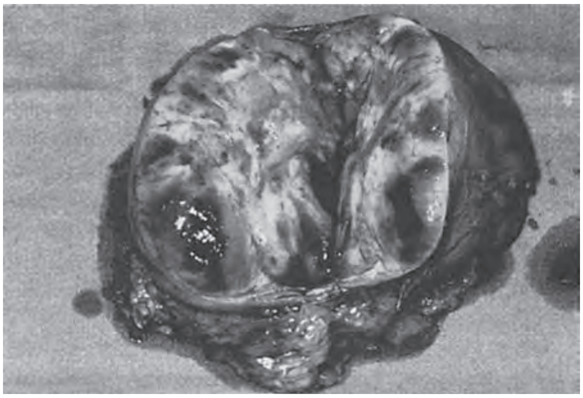


Figure 20.20 Gross specimen of phaeochromocytoma diagnosed during pregnancy and excised at 36 weeks synchronously with caesarean section delivery. The tumour shows typical haemorrhagic appearances on the cut surface.

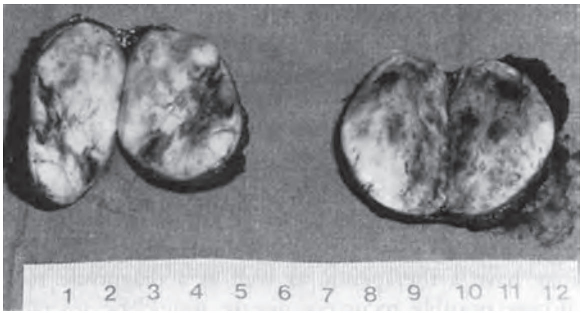


Figure 20.21 Gross specimen of bilateral phaeochromocytoma excised from a 5-year-old boy.

Phaeochromocytoma in children

When phaeochromocytoma occurs in children there is an increased incidence of bilateral, multiple and extra-adrenal tumours because of the association with MEN 2 syndromes (Figure 20.21). In contrast to adults the tumours are less commonly malignant and the hypertension more often sustained. Hypertension in children always demands the fullest investigation and surgical resection of a phaeochromocytoma is usually followed by an excellent clinical result.

Phaeochromocytoma and multiple endocrine neoplasia 2A and 2B

MEN is a syndrome characterized by medullary thyroid carcinoma and phaeochromocytoma. In the MEN 2A variant, primary hyperparathyroidism is also associated, whereas in MEN 2B patients have characteristic facies, Marfinoid habitus, skeletal abnormalities and mucocutaneous ganglioneuromas. Hyperparathyroidism does not occur in MEN 2B.

Germ-line mutations in the *Ret* proto-oncogene are responsible for MEN 2A and 2B. The gene is inherited in an autosomal dominant fashion with high penetrance and variable expression. All patients with MEN 2 will develop medullary

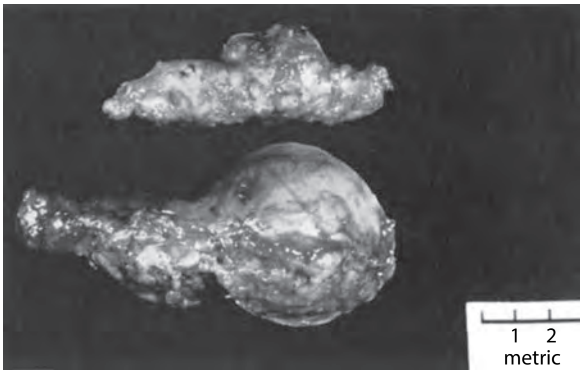


Figure 20.22 Gross specimen of bilateral phaeochromocytomas excised from a patient who had previously undergone total thyroidectomy for medullary thyroid carcinoma as part of the multiple endocrine neoplasia 2B syndrome.

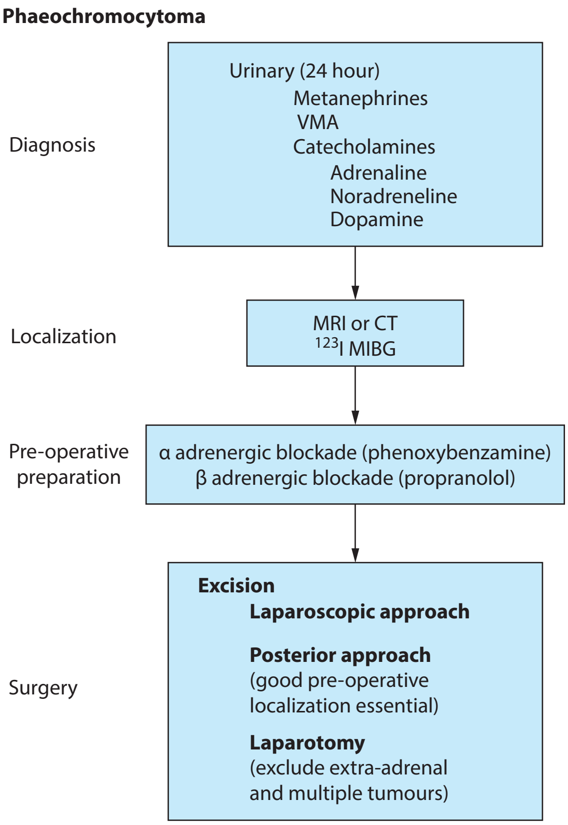

Phaeochromocytoma
Diagnosis
Urinary (24 hour)
Metanephrines
VMA
Catecholamines
Adrenaline
Noradreneline
Dopamine
Localization
MRI or CT
^{123}I MIBG
Pre-operative preparation
α adrenergic blockade (phenoxybenzamine)
β adrenergic blockade (propranolol)
Surgery
Excision
Laparoscopic approach
Posterior approach
(good pre-operative localization essential)
Laparotomy
(exclude extra-adrenal and multiple tumours)

Figure 20.23 Investigation and treatment of phaeochromocytoma.

thyroid cancer but development of phaeochromocytoma is variable.

When present adrenal disease in MEN 2 is bilateral, passing through a phase of hyperplasia to nodularity and multiple phaeochromocytomas. Once the disease has been confirmed through measurement of urinary catecholamines and their metabolites, bilateral adrenalectomy should be performed. Some advocate unilateral adrenalectomy when the disease appears localized to one side but this approach remains controversial as the natural history of the disease clearly indicates that bilateral disease eventually develops in all affected individuals (Figure 20.22). After adrenalectomy careful follow-up is mandatory in all patients, but after unilateral adrenalectomy this is especially pertinent in order to monitor the development of disease in the contralateral gland.

Summary

The investigation and treatment of phaeochromocytoma is summarized in Figure 20.23.

Other adrenal tumours

Adrenal incidentaloma

Increasing use of sophisticated imaging modalities such as ultrasound, CT and MRI, and the ever-increasing resolution that both software and hardware are achieving, has meant that the number of incidentally discovered adrenal masses has increased over the past 10 years. Silent adrenal masses are now identified in 0.5–4.5% of patients undergoing CT for reasons other than suspected adrenal pathology, a figure approaching the incidence of incidentalomas identified at postmortem (1.5–5.7%). Adrenal incidentalomas present the clinician with a diagnostic problem, particularly in excluding malignancy tumour but also with respect to functioning potential.

The essential step in the investigation of an incidentally discovered adrenal tumour is to establish whether the lesion is functioning. Biochemical screening should include measurement of electrolytes, 24 hour urinary VMA and metanephrine levels, and 09.00 and midnight serum cortisol and aldosterone. The 1 mg overnight dexamethasone suppression test should exclude Cushing syndrome.

The majority of incidentalomas (35–95%) are benign, non-functioning adrenal adenomas. They are more prevalent in older women and in obese, diabetic patients. Tumours are classically smooth, round and <3 cm in size. They are homogeneous and enhance only minimally after contrast injection on CT. With MRI, signal intensity is the same as normal adrenal tissue; this is helpful in distinguishing secondary metastatic lesions in the adrenal, which tend to have intense T_2 weighted images (Figure 20.24). Calcification, haemorrhage and necrosis are uncommon.

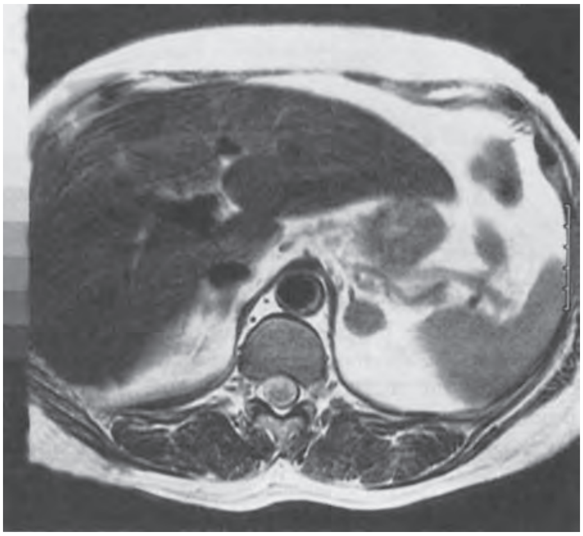


Figure 20.24 MRI scan of a non-functioning left adrenal incidentaloma.

In-/out-of-phase MRI scanning is useful in characterising the fat content of adrenal incidental adenomas and identifying them as benign. In rare instances in which tumours are not characterized, then positron emission tomography scanning may be useful in larger (>4 cm) tumours.

Cytology may aid in the diagnosis of malignancy. Following the exclusion of a phaeochromocytoma, fine needle aspiration cytology can be selectively and carefully employed if a secondary adrenal malignant deposit is suspected.

Adrenal cysts also cause diagnostic confusion. These are usually small and symptomless but when large may present as a tumour mass displacing the kidney. Drainage under ultrasound or CT guidance is recommended and the fluid sent for cytological analysis. Adenomyelolipomas are benign, non-functioning tumours which have a characteristic appearance on MRI and only require excision if there are significant symptoms.

The incidence of malignancy in adrenal tumours increases with size of lesion and for this reason surgical resection is recommended for all lesions >4 cm. MRI or CT scans demonstrating non-homogeneity may also influence the decision to perform surgery.

Smaller lesions can be monitored by interval CT, MRI or ultrasound scans and excision considered when there is an increase in size. Surgical excision should also be performed in patients under 50 years of age because of the increased malignancy risk for adrenal lesions in the younger subject.

Secondary adrenal tumours

Many malignant neoplasms metastasize to the adrenal glands. Among the more common tumours are breast, bronchus and melanoma. Adrenocortical hormone production may be reduced when large metastases cause significant adrenal destruction resulting in an acute Addisonian crisis.

Neuroblastoma

Presentation

These tumours of neural crest origin and occur mainly in children, with over 60% presenting in the first year of life. They occur in the adrenal medulla, adjacent retroperitoneal tissue and along the sympathetic ganglia. The majority of the tumours occur in the abdomen (75%), the remainder occurring in the thorax (20%) and neck (5%). Aggressive malignancies, they invade adjacent local structures such as kidney, spleen, liver and pancreas. Metastatic spread occurs early via the bloodstream and lymphatics and is frequently present at initial presentation. Two distinct patterns of metastatic spread are recognized: Hutchinson's type (tumour on the left side producing metastases to the orbit, skull and long bones) and Pepper's type (tumour on the right side with metastatic spread to the liver).

Approximately 50% of children with this tumour present with a large, symptomless abdominal mass. The other 50% present with symptoms including anorexia, nausea, vomiting and diarrhoea (tumours producing vasoactive intestinal peptide). Over 90% of neuroblastomas produce catecholamines and hypertension and/or flushing may be a feature. A 24 hour urine collection for VMA, metanephrines and catecholamines is therefore mandatory in these patients. Dumb-bell tumours (tumours extending into the spinal canal) may produce neurological symptoms. The main differential diagnosis of neuroblastoma is nephroblastoma (Wilms' tumour).

Tumour staging

Tumours are staged with a combination of chest radiograph, CT/MRI scanning and skeletal survey. MIBG scintigraphy may be useful in identifying primary tumours as well as residual or recurrent tumour:

- Stage I: tumours are confined to the adrenal gland and are totally excised.
- Stage II: tumours extend beyond the organ of origin but do not cross the midline; ipsilateral lymph nodes may be involved.
- Stage III: tumours cross the midline.
- Stage IV: distant metastatic spread present.

Treatment

Surgical excision remains the mainstay of treatment. When complete excision is not possible, tumour debulking followed by either radiotherapy and/or combination chemotherapy with vincristine and cyclophosphamide may be of benefit. Radiotherapy is of particular benefit in painful metastatic deposits.

Prognosis

Prognosis is dependent on age and tumour stage. Children aged <2 years have a better 2 year survival rate than their older counterparts (77% compared with 38%). The 2 year survival rate worsens with stage: stage I (100%), stage II (82%), stage III (42%) and stage IV (30%). Factors improving stage-specific prognosis include a high urinary VMA to homovanillic acid ratio and low serum neurone-specific enolase level.

Adrenalectomy

Background

The first description of an adrenalectomy was provided by Thornton in 1890. He employed an approach to the right adrenal gland using an incision for cholecystectomy, previously described by Carl von Langenbüch of Berlin in 1882. For many years the adrenal glands were approached via incisions described for renal surgery. Unfortunately these incisions were frequently too low to gain adequate exposure and surgeons began to site incisions progressively higher. In 1932 Lennox Broster of London (a pioneer of adrenal surgery) described a posterior, intercostal, transpleural adrenalectomy. He had performed a laparotomy a few weeks previously in the same patient to inspect the adrenals and look for any extra-adrenal tissue. In 1936 Hugh Young of Baltimore, MD, devised a simultaneous, bilateral, posterior approach, enabling him to inspect both adrenal glands, a technique associated with little postoperative morbidity. Prior to the advent of CT scanning, exploratory laparotomy played a pivotal role in both localizing adrenal and extra-adrenal tumours and providing access for adrenalectomy. With modern-day localizing techniques, however, this approach has become virtually obsolete.

The main approaches to the adrenal glands are anterior and posterior or posterolateral. The description by Gagner in 1992 of an endoscopic technique has provided surgeons with the additional option of performing adrenalectomy as either an 'open' or a 'minimally invasive' (endoscopic) procedure. Each method of adrenalectomy has both advantages and disadvantages and the approach of choice is dependent on a number of factors including tumour pathology, tumour size, patient physique, previous surgery and the personal preference of the surgeon. Despite the various options, adrenalectomy still provides the modern-day endocrine surgeon with a significant surgical challenge.

Open anterior approach

With improved localization techniques the *open anterior approach* tends to be reserved for the removal of large (<6–8 cm), malignant or potentially malignant tumours or in patients in whom localizing procedures have been equivocal. It provides excellent exposure of one or both adrenal glands and enables examination of the abdominal viscera and sympathetic chain to search for extra-adrenal tissue. The patient is placed supine on the operating table and a transverse, subcostal incision made. This may either be unilateral or be extended across the midline when exposure of both adrenal glands is required. Alternatively a vertical midline incision may be preferred. The major disadvantages of this approach are:

- in patients with Cushing syndrome, when wound complications including infection and dehiscence contribute significantly to morbidity
- it necessitates entering the peritoneal cavity producing a postoperative ileus
- there is an increased incidence of atelectasis, pulmonary collapse and chest infection.

Laparoscopic 'anterior' approach

Laparoscopic adrenalectomy can be performed effectively and safely. This approach was quickly popularized and subsequent reports confirmed its feasibility. Most surgeons currently use a lateral flank approach with the patient lying with the side to be operated on uppermost. A pneumoperitoneum is created and three or, when needed, four laparoscopic ports are inserted below the costal margin. Ports should be either 5–12 mm or 5 mm in size depending on the kit available to the surgeon and personal choice. Placing all 5–12 mm ports enables the camera to be moved from port to port to improve visualization. A significant advantage of the flank approach is that once the spleen is mobilized on the left side it acts as a retractor under its own weight. Similarly the liver on the right requires minimal retraction.

The anterior laparoscopic approach allows excellent visualization of the adrenal glands and the vascular pedicles both sides, particularly the adrenal veins. It is an ideal approach for Conn's tumours and smaller phaeochromocytomas (<6 cm) as the adrenal vein can be divided early in the procedure. Operative times have dramatically improved with experience and new instrumentation to an extent that bilateral laparoscopic adrenalectomy is now readily performed.

Right-sided laparoscopic adrenalectomy

Four subcostal ports are placed, with the most medial port being a 5 mm port to allow liver retraction, and the other three 5–12 mm ports. The liver is elevated and the inferior vena cava identified. The peritoneal reflection of the liver onto the back wall is incised and the liver elevated further. The peritoneum alongside the cava is incised revealing the adrenal gland. The dissection proceeds carefully cephalad alongside the cava until the adrenal vein is identified. This is gently mobilized and then taken between clips. Once the adrenal vein is divided the adrenal gland can be elevated off the posterior abdominal wall and freed either using a hook or with a device such as the Harmonic Scalpel or LigaSure. The adrenal gland is dissected free usually in a medial to lateral direction. Normally there are no obviously large individual arteries to be divided between clips but this can change in the case of large tumours, particularly phaeochromocytoma (which may on occasion be very vascular). Care must be taken not to place dissecting graspers through the capsule of the adrenal gland or to tear the liver capsule by vigorous retraction too early in the procedure as both events often cause troublesome bleeding that hinders visualization.

The gland is placed into a retrieval bag and removed via the lateral port. Wounds are closed with a subcuticular stitch.

Left-sided laparoscopic adrenalectomy

Three ports are usually sufficient to allow successful mobilization and removal of the left adrenal gland. The most medial port may be a 5 mm port, the other two 5–12 mm. Care should be take not to place the ports too medially as, if a fourth port is needed, space must be left for this to be placed without causing problems. The splenic flexure of the colon should be fully mobilized inferiorly to allow identification of the renal vein. The lienorenal ligament is divided and the spleen mobilized medially. Gerota's fascia is divided allowing the spleen to be further mobilized medially. Dividing the fascia allows identification of the adrenal gland. Looking for the gland too early (prior to dividing Gerota's fascia) is a cause of surgeons struggling to find a left gland. In this situation it is possible to mistake the tail of the pancreas for the adrenal, increasing potential morbidity.

Once identified (usually running cephalad alongside the medial aspect of the kidney, not superior to it) the gland it traced inferiorly to identify the adrenal vein, usually found in the 6 o'clock position relative to the gland. As on the right side, division of the vein between clips allows the adrenal to be elevated off the posterior abdominal wall and dissected free. The gland, once freed, is placed in a retrieval bag and removed via the lateral port.

Right-sided tumours vary more than those on the left side in terms of difficulty. Identification of the tumour is much easier on the right than on the left. With small tumours, identifying and dividing the adrenal vein is often very straightforward. However, a large right-sided tumour extending behind the cava and under the liver, often with very short adrenal vein, can become an extremely difficult procedure, even for very experienced adrenal surgeons. On the left side, in obese patients identification of the gland can sometimes prove very difficult. When the gland cannot be located, it is likely to be found more medially and inferiorly than expected. Tracing the renal vein to find the adrenal vein can sometimes be useful and indeed many surgeons teach this technique routinely. Failing to fully mobilize the spleen is usually the cause of access problems on the left side; on occasion it may be necessary to place a port to allow the spleen to be retracted.

Postoperatively patients can be mobilized quickly and inpatient stay is significantly reduced compared with open adrenalectomy. Perhaps the major advantage of laparoscopic adrenalectomy is seen in patients with Cushing syndrome. Wound size is markedly reduced in these patients and consequently the wound-associated morbidity has dramatically improved.

Anterior laparoscopic adrenalectomy is unsuitable for adrenal tumours >8 cm and for malignant and potentially malignant tumours. A relative contraindication is previous abdominal surgery. Identification of smaller left-sided tumours may prove difficult.

Open posterior approach

This approach has now been largely abandoned owing to the poor view and the morbidity and pain associated. The posterior approach to the adrenal gland was originally described by Hugh Young in 1936. Most endocrine surgeons used to favour the Turner–Warwick modification of the posterior approach, in which the eleventh or twelfth ribs are removed extrapleurally. This incision may also be slightly modified by making a 'hockey stick'-shaped incision that initially runs parallel to the spinal column.

The main disadvantages of this approach are restricted visualization and access, particularly on the right side where the right adrenal vein is short and may be difficult to secure.

Posterior endoscopic approach

The posterior endoscopic approach to the adrenal glands has now become an established approach. It has been popularized by Professor Martin Walz, Essen, Germany. It is suitable for smaller tumours (<4 cm), particularly Conn's tumours, patients needing bilateral adrenalectomy (failed treatment of pituitary Cushing disease, bilateral phaeochromocytoma) as there is no need to turn the patient from one side to the other and in patients in whom previous abdominal surgery may make the anterior approach difficult or hazardous. Patients are positioned as for an open posterior approach and a balloon is inserted and inflated in the retroperitoneal space below the twelfth rib. Ports are inserted into the artificial space and the adrenal gland is visualized. The major drawbacks with this approach are:

- difficult anatomical orientation compared with the anterior laparoscopic approach
- very limited operating space
- normal endoscopic instruments may be too long
- even minor bleeding may cause significant visualization problems.

Because of these problems the approach has not initially proved popular with surgeons. The approach, however, does offer an alternative technique in patients with suitable adrenal tumours and many endocrine surgeons, having gained experience in the anterior laparoscopic approach, are now learning this technique to further enhance their range of operative options.

GUIDE TO FURTHER READING

Abel JJ, Crawford AC. On the blood pressure raising constituents of the supra renal capsule. *Johns Hopkins Hosp Bull* 1897;**8**:151.

Addison T. *On the Constitutional and Local Effects of Disease of the Supra Renal Capsules.* London, UK: Samuel Highley, 1855.

Amador E. Adrenal hemorrhage during anticoagulant therapy. A clinical and pathological study often cases. *Ann Intern Med* 1965;**63**:559–571.

Armstrong MD, McMillan A, Shaw KNF. 3-Methoxy-4-hydroxy-D-mandellic acid, a urinary metabolite of norepinephrine. *Biochim Biophys Acta* 1957;**25**:422–428.

Auda SP, Brennan MF, Gill JR Jr. Evolution of the surgical management of primary aldosteronism. *Ann Surg* 1980;**191**:1–7.

Azzopardi JG, Williams ED. Pathology of 'nonendocrine' tumors associated with Cushing's syndrome. *Cancer* 1968;**22**:274–286.

Barbeau A, Marc-Aurele J, Brouillet J, *et al.* Roux: Thesis Lausanne. Le pheochromocytome bilateral: presentation d'un cas et revue de la litterature. *J Union Med Can* 1958;**87**:165–172.

Biglieri EG, Irony I, Kater CE. Identification and implications of new types of mineralocorticoid hypertension. *J Steroid Biochem* 1989;**32**(1B):199–204.

Boggan JE, Tyrrell JB, Wilson CB. Transsphenoidal microsurgical management of Cushing's disease. Report of 100 cases. *J Neurosurg* 1983;**59**:195–200.

Boutros AR, Bravo EL, Zanettin G, Straffon RA. Perioperative management of 63 patients with pheochromocytoma. *Cleveland Clin J Med* 1990;**57**:613–617.

Bravo EL. Physiology of the adrenal cortex. *Urol Clin North Am* 1989;**16**:433–437.

Bravo EL. Evolving concepts in the pathophysiology, diagnosis, and treatment of pheochromocytoma. *Endocr Rev* 1994;**15**:356–368.

Broster LR, Hill HG, Greenfield JG. Adrenogenital syndrome and unilateral adrenalectomy. *Am J Surg* 1932;**19**:557–570.

Brown WH. A case of pleuriglandular syndrome: diabetes of bearded woman. *Lancet* 1928;**ii**:1022–1027.

Brown-Séquard C. Recherches experimentales sur la physiologie et la pathologie des capsules surrenales. *Arch Gen Med (Paris)* 1856;**8**:385.

Burke CW, Beardwell CG. Cushing's syndrome. An evaluation of the clinical usefulness of urinary free cortisol and other urinary steroid measurements in diagnosis. *Q J Med* 1973;**42**:175–204.

Burke CW, Adams CB, Esiri MM, *et al.* Transsphenoidal surgery for Cushing's disease: does what is removed determine the endocrine outcome? *Clin Endocrinol* 1990;**33**:525–537.

Carney JA, Sizemore GW, Tyce GM. Bilateral adrenal medullary hyperplasia in multiple endocrine neoplasia, type 2: the precursor of bilateral pheochromocytoma. *Mayo Clin Proc* 1975;**50**:3–10.

Chang A, Glazer HS, Lee JK, *et al.* Adrenal gland: MR imaging. *Radiology.* 1987;**163**:123–128.

Cirillo RL Jr, Bennett WF, Vitellas KM, *et al.* Pathology of the adrenal gland: imaging features. *AJR Am J Roentgenol* 1998;**170**:429–435.

Conn JW. Part I. Painting background. Part II. Primary aldosteronism, a new clinical syndrome. *Lab Clin Med* 1955;**45**:3.

Crapo L. Cushing's syndrome: a review of diagnostic tests. *Metab Clin Exp* 1979;**28**:955–977.

Cushing H. The pituitary body and its disorders. *Johns Hopkins Hosp Bull* 1932;**50**:137.

Cutroneo KR, Rokowski R, Counts DF. Glucocorticoids and collagen synthesis: comparison of *in vivo* and cell culture studies. *Collagen Relate Res* 1981;**1**:557–568.

Daly PA, Landsberg L. Phaeochromocytoma: diagnosis and management. *Baillières Clin Endocrinol Metab* 1992;**6**:143–166.

Doppman JL, Oldfield E, Krudy AG, *et al.* Petrosal sinus sampling for Cushing syndrome: anatomical and technical considerations. *Radiology* 1984;**150**:99–103.

Duh QY, Siperstein AE, Clark OH, *et al.* Laparoscopic adrenalectomy. Comparison of the lateral and posterior approaches. *Arch Surg* 1996;**131**:870–875; discussion 875–876.

Dunnick NR, Heaston D, Halvorsen R, *et al.* CT appearance of adrenal cortical carcinoma. *J Comput Assist Tomogr* 1982;**6**:978–982.

Dwyer AJ, Frank JA, Doppman JL, *et al.* Pituitary adenomas in patients with Cushing's disease: initial experience with Gd-DTPA-enhanced MRI imaging. *Radiology* 1987;**163**:421.

Edis AJ, Grant CS, Egdahal RH (eds). *Manual of Endocrine Surgery,* 2nd edn. New York, NY: Springer, 1984.

Edwards CR, Besser GM. Diseases of the hypothalamus and pituitary gland. *Clin Endocrinol Metab* 1974;**3**:475–505.

Evans AE, D'Angio GJ, Randolph J. A proposed staging for children with neuroblastoma. Children's Cancer Study Group A. *Cancer* 1971;**27**:374–378.

Farwell AP, Devlin JT, Stewart JA. Total suppression of cortisol excretion by ketoconazole in the therapy of the ectopic adrenocorticotropic hormone syndrome. *Am J Med* 1988;**84**:1063–1066.

Favio G, Lumachi FN. Cushing's syndrome. In: Clark OH, Duh QY (eds) *Textbook of Endocrine Surgery.* Philadelphia, PA: W.B. Saunders, 1997.

Ferriss JB, Brown JJ, Fraser R, *et al.* Results of adrenal surgery in patients with hypertension, aldosterone excess, and low plasma renin concentration. *BMJ* 1975;**1**(5950):135–138.

Finkelstein M, Shaefer JM. Inborn errors of steroid biosynthesis. *Physiol Rev* 1979;**59**:353–406.

Fishman LN, Liddle GW, Island DP, *et al.* Effects of aminogluthamide on adrenal function in man. *J Clin Endocrinol Metab* 1987;**27**:481–488.

Frankel F. Ein Fall von doppelseitigem, vollig latent verlaufen Nebennierentumor und zleichzeitizer Nephritis. *Vichows Arch Pathol Anat Klin Med* 1886;**103**:244–263.

Friesen SR, Thompson NW (eds). *Surgical Endocrinology.* Philadelphia, PA: Lippincott, 1990.

Fudge TL, McKinnon WM, Geary WL. Current surgical management of pheochromocytoma during pregnancy. *Arch Surg* 1980;**115**:1224–1225.

Gagner M, Lacroix A, Bolte E. Laparoscopic adrenalectomy in Cushing's syndrome and pheochromocytoma. *N Engl J Med* 1992;**327**:1033.

Granberg PO, Adamson U, Cohn KH, *et al.* The management of patients with primary aldosteronism. *World J Surg* 1982;**6**:757–764.

Grant CS, Carpenter P, van Heerden JA, Hamberger B. Primary aldosteronism. Clinical management. *Arch Surg* 1984;**119**:585–590.

Grant CS, Carney JA, Carpenter PC, van Heerden JA. Primary pigmented nodular adrenocortical disease: diagnosis and management. *Surgery* 1986;**100**:1178–1184.

Gross MD, Wilton GP, Shapiro B, *et al.* Functional and scintigraphic evaluation of the silent adrenal mass. *J Nucl Med* 1987;**28**:1401–1407.

Hall R. Addison's disease. *Med Int* 1981;**6**:276–278.

Henley DJ, van Heerden JA, Grant CS, *et al.* Adrenal cortical carcinoma – a continuing challenge. *Surgery* 1983;**94**:926–931.

Herrera MF, Grant CS, van Heerden JA, *et al.* Incidentally discovered adrenal tumors: an institutional perspective. *Surgery* 1991;**110**:1014–1021.

Horgan S, Sinanan M, Helton WS, Pellegrini CA. Use of laparoscopic techniques improves outcome from adrenalectomy. *Am J Surg* 1997;**173**:371–374.

Hughes IA. Congenital and acquired disorders of the adrenal cortex. *Clin Endocrinol Metab* 1982;**11**:125–131.

Hughes IA, Read GF. Simultaneous plasma and saliva steroid measurements as an index of control in congenital adrenal hyperplasia (CAH). A longitudinal study. *Horm Res* 1982;**16**:142–150.

Hughes IA, Winter JS. The relationships between serum concentrations of 17OH-progesterone and other serum and urinary steroids in patients with congenital adrenal hyperplasia. *J Clin Endocrinol Metab* 1978;**46**:98–104.

Hughes IA, Riad-Fahmy D, Griffiths K. Plasma 17 OH-progesterone concentrations in newborn infants. *Arch Dis Child* 1979;**54**:347–349.

Hunt TK, Schambelan M, Biglieri EG. Selection of patients and operative approach in primary aldosteronism. *Ann Surg* 1975;**182**:353–361.

Javadpour N, Woltering EA, Brennan MF. Adrenal neoplasms. *Curr Prob Surg* 1980;**17**:1–52.

Jeffcoate WJ, Rees LH, Tomlin S, *et al.* Metyrapone in long-term management of Cushing's disease. *BMJ* 1977;**ii**:215–217.

Jimenez C, Cabanillas ME, Santarpia L, *et al.* Use of the tyrosine kinase inhibitor sunitinib in a patient with von Hippel-Lindau disease: targeting angiogenic factors in pheochromocytoma and other von Hippel-Lindau disease-related tumors. *J Clin Endocrinol Metab* 2009;**94**:386–391.

Johnston RR, Eger EI II, Wilson C. A comparative interaction of epinephrine with enflurane, isoflurane, and halothane in man. *Anesth Analg* 1979;**55**:709–712.

Joshua AM, Ezzat S, Asa SL, *et al.* Rationale and evidence for sunitinib in the treatment of malignant paraganglioma/pheochromocytoma. *J Clin Endocrinol Metab* 2009;**94**:5–9.

Landsberg L, Young JB. Catecholamines and the adrenal medulla. In: Bondy PK, Rosenburg LE (eds) *Metabolic Control and Disease.* Philadelphia, PA: W.B. Saunders, 1980.

Liddle GW. Test of pituitary-adrenal suppressibility in the diagnosis of Cushing's syndrome. *J Clin Endocrinol Metab* 1960;**20**:1539.

Liddle GW, Island D, Meador CK. Normal and abnormal regulation of corticotrophin secretion. *Recent Prog Horm Res* 1962;**15**:125–166.

Manger WM, Gifford RW Jr. *Pheochromocytoma.* New York, NY: Springer, 1977.

Marescaux J, Mutter D, Wheeler MH. Laparoscopic right and left adrenalectomies. *Surg Endosc* 1996;**10**:912–915.

Mayo CH. Paroxysmal hypertension with tumour of retroperitoneal nerve. *JAMA* 1927;**89**:1047.

Minno AM, Bennett WA, Kvale WF. Phaeochromocytoma. *N Engl J Med* 1954;**251**:959–965.

Montgomery DAD, Welbourn RB. Cushing's syndrome: 20 years after adrenalectomy. *Br J Surg* 1978;**65**:210–220.

Mukherjee JJ, Peppercorn PD, Reznek RH, *et al.* Pheochromocytoma: effect of nonionic contrast medium in CT on circulating catecholamine levels. *Radiology* 1997;**202**:227–231.

Mulligan LM, Kwok JB Healey CS, *et al.* Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* 1993;**363**:458–460.

Navaratnarajah M, White DC. Labetalol and phaeochromocytoma. *Br J Anaesth* 1984;**56**:1179.

Nelson DH, Meakin JW, Thorn GW. ACTH producing pituitary tumours following adrenalectomy for Cushing's syndrome. *Ann Intern Med* 1960;**52**:560.

Nerup J. Addison's disease – clinical studies. A report of 108 cases. *Acta Endocrinol* 1974;**76**:127–141.

Neumann PJ. Addison's disease and acute adrenal haemorrhage. In: Clark OH, Duh QY (eds) *Textbook of Endocrine Surgery.* Philadelphia, PA: W.B. Saunders, 1997.

Nicholson JP Jr, Vaughn ED Jr, Pickering TG, *et al.* Pheochromocytoma and prazosin. *Ann Intern Med* 1983;**99**:477–479.

Obara T, Ito Y, Okamoto T, *et al.* Risk factors associated with postoperative persistent hypertension in patients with primary aldosteronism. *Surgery* 1992;**112**:987–993.

Oldfield EH, Chrousos GP, Schulte HM, *et al.* Preoperative lateralization of ACTH-secreting pituitary microadenomas by bilateral and simultaneous inferior petrosal venous sinus sampling. *N Engl J Med* 1985;**312**:100–103.

Oliver G, Sharpey-Shafer EA. The physiological effects of extracts of the supra renal capsule. *J Physiol (Lond)* 1895;**18**:230.

Orth DN. Metyrapone is useful only as adjunctive therapy in Cushing's disease. *Ann Intern Med* 1978;**89**:128–130.

Padberg BC, Holl K, Schroder S. Pathology of multiple endocrine neoplasias 2A and 2B: a review. *Horm Res* 1992;**38**(Suppl. 2):24–30.

Pavlatos FC, Smilo RP, Fordham PH. A rapid screening test for Cushing's syndrome. *JAMA* 1965;**193**:720.

Peppercorn PD, Grossman AB, Reznek RH. Imaging of incidentally discovered adrenal masses. *Clin Endocrinol* 1998;**48**:379–388.

Pick L. Das Ganglioma embryonale sympathicum. *Berlin Klin Wochenschr* 1912;**19**:16–22.

Plumpton FS, Besser GM, Cole PV. Corticosteroid treatment and surgery. 2. The management of steroid cover. *Anaesthesia* 1969;**24**:12–18.

Pommier RF, Brennan MF. An eleven-year experience with adrenocortical carcinoma. *Surgery* 1992;**112**:963–970; discussion 970–971.

Pont A, Williams PL, Loose DS, *et al.* Ketoconazole blocks adrenal steroid synthesis. *Ann Intern Med* 1982;**97**:370–372.

Proye CA, Huart JY, Cuvillier XD, *et al.* Safety of the posterior approach in adrenal surgery: experience in 105 cases. *Surgery* 1993;**114**:1126–1131.

Quinn SJ, Williams GH. Regulation of aldosterone secretion. *Annu Rev Physiol* 1988;**50**:409–426.

Remine WH, Chong GC, Van Heerden JA, *et al.* Current management of pheochromocytoma. *Ann Surg* 1974;**179**:740–748.

Rolleston HD. *The Endocrine Organs in Health and Disease.* Oxford, UK: Oxford University Press, 1936.

Rovit RL, Duane TD. Cushing's syndrome and pituitary tumors. Pathophysiology and ocular manifestations of ACTH-secreting pituitary adenomas. *Am J Med* 1969;**46**:416–427.

Ruder HJ, Loriaux DL, Lipsett MB. Severe osteopenia in young adults associated with Cushing's syndrome due to micronodular adrenal disease. *J Clin Endocrinol Metab* 1974;**39**:1138–1147.

Savage MO. Congenital adrenal hyperplasia. *Clin Endocrinol Metab* 1985;**14**:893–909.

Schenker JG, Chowers I. Pheochromocytoma and pregnancy. Review of 89 cases. *Obstet Gynecol Surv* 1971;**26**:739–747.

Schenker JG, Granat M. Phaeochromocytoma and pregnancy – an updated appraisal. *Aust NZ J Obstet Gynaecol* 1982;**22**:1–11.

Schteingart DE, Seabold JE, Gross MD, Swanson DP. Iodocholesterol adrenal tissue uptake and imaging adrenal neoplasms. *J Clin Endocrinol Metab* 1981;**52**:1156–1161.

Schteingart DE, Motazedi A, Noonan RA, Thompson NW. Treatment of adrenal carcinomas. *Arch Surg* 1982;**117**:1142–1146.

Scott HW Jr, Dean RH, Lea JW, *et al.* Surgical experience with retrogastric and retropancreatic pheochromocytomas. *Surgery* 1982;**92**:853–865.

Serfas D, Shoback DM, Lorell BH. Phaeochromocytoma and hypertrophic cardiomyopathy: apparent suppression of symptoms and noradrenaline secretion by calcium-channel blockade. *Lancet* 1983;**ii**:711–713.

Soreide JA, Brabrand K, Thoresen SO. Adrenal cortical carcinoma in Norway, 1970–84. *World J Surg* 1992;**16**:663–667.

Staren ED, Prinz RA. Adrenalectomy in the era of laparoscopy. *Surgery* 1996;**120**:706–711.

Tam PK. Paediatric solid tumours. In: Morris PJ, Malt RA (eds) *Oxford Textbook of Surgery.* New York, NY: Oxford University Press, 1994.

Thompson NW, Cheung PS. Diagnosis and treatment of functioning and nonfunctioning *adrenocortical* neoplasms including incidentalomas. *Surg Clin North Am* 1987;**67**:423–436.

Thompson GB, Grant CS, van Heerden JA, *et al.* Laparoscopic versus open posterior adrenalectomy: a case control study of 100 patients. *Surgery* 1997;**122**:1136.

Thornton JK. Abdominal nephrectomy for large sarcoma of the left suprarenal capsule. *Clin Soc Trans (Lond)* 1890;**23**:150–153.

Tibblin S, Dymling JF, Ingemansson S, Telenius-Berg M. Unilateral versus bilateral adrenalectomy in multiple endocrine neoplasia IIA. *World J Surg* 1983;**7**:201–208.

Troncone L, Rufini V, Montemaggi P, *et al.* The diagnostic and therapeutic utility of radioiodinated metaiodobenzylguanidine (MIBG). 5 years of experience. *Eur J Nucl Med* 1990;**16**:325–335.

Turner-Warwick RT. The supra costal approach to the renal area. *Br J Urol* 1965;**37**:671–676.

van Heerden JA, Sheps SG, Hamberger B, *et al.* Pheochromocytoma: current status and changing trends. *Surgery* 1982;**91**:367–373.

van Heerden JA, Grant CS, Weaver AC. Primary carcinoma of the adrenal cortex: an institutional surgical perspective. *Acta Chir Austrica* 1993;**25**:216.

Veldhuis JD, Iranmanesh A, Lizarralde G, Johnson ML. Amplitude modulation of a burstlike mode of cortisol secretion subserves the circadian glucocorticoid rhythm. *Am J Physiol* 1989;**257**(1 Pt 1):E6–14.

Verner JV, Morrison AB. Endocrine pancreatic islet disease with diarrhea. Report of a case due to diffuse hyperplasia of nonbeta islet tissue with a review of 54 additional cases. *Arch Intern Med* 1974;**133**:492–499.

von Langenbuch C. Ein Fall von Exstirpation der Gallenblase. *Berlin Klin Wochenschr* 1882;**19**:725–727.

Walz MK, Peitgen K, Hoermann R, *et al.* Posterior retroperitoneoscopy as a new minimally invasive approach for adrenalectomy: results of 30 adrenalectomies in 27 patients. *World J Surg* 1996;**20**:769–774.

Welbourn RB (ed.). *The History of Endocrine Surgery.* New York, NY: Praeger, 1990.

Welbourn RB, Manolas KJ. *Endocrine Surgery.* London, UK: Butterworth, 1983.

Werbel SS, Ober KP. Acute adrenal insufficiency. *Endocrinol Metab Clin North Am* 1993;**22**:303–328.

White FE, White MC, Drury PL, *et al.* Value of computed tomography of the abdomen and chest in investigation of Cushing's syndrome. *BMJ Clin Res Ed* 1982;**284**:771–774.

Young HG. A technique for simultaneous exposure and operation on the adrenals. *Surg Gynecol Obstet* 1936;**54**:179.

Young WF Jr. Pheochromocytoma and primary aldosteronism: diagnostic approaches. *Endocrinol Metab Clin North Am* 1997;**26**:801–827.

Young WF Jr, Klee GG. Primary aldosteronism. Diagnostic evaluation. *Endocrinol Metab Clin North Am* 1988;**17**:367–395.

Young AE, Smellie WJB. The adrenal glands. In: Fardon JR (ed.) *Breast and Endocrine Surgery.* London, UK: W.B. Saunders, 1997.

Websites

http://www.acr.org/media/ACR/documents/app_criteria/IncidentallyDiscoveredAdrenalMassDoc7.aspx

http://www.websurg.com/ref/Laparoscopic_left_adrenalectomy_for_Conns_disease%253A_transabdominal_approach-vd01en1409e.htm

http://www.websurg.com/ref/Laparoscopic_retroperitoneal_left_adrenalectomy-vd01en1795.htm

http://www.websurg.com/ref/Laparoscopic_retroperitoneal_right_adrenalectomy_in_an_obese_patient-vd01en1798.htm

http://www.websurg.com/ref/Laparoscopic_right_adrenalectomy-vd01en1801.htm

CHAPTER 21

Disorders of the abdominal wall, peritoneal cavity and retroperitoneum

SIR ALFRED CUSCHIERI

Surgical anatomy

The two terms abdomen (or abdominal cavity) and peritoneal cavity are not strictly synonymous. The abdomen (or abdominal cavity) refers to the musculoaponeurotic and bony walls that enclose a region lined on the inside by the peritoneum. By contrast, the peritoneal cavity denotes the space enclosed by the peritoneal lining and contains some but not all of the abdominal viscera. The retroperitoneal space lies behind the posterior wall of the peritoneal cavity and contains adipose and areolar 'packing tissue' in which lie the retroperitoneal organs. This is continuous with the rest of the extraperitoneal space and is filled with the same packing tissue separating the musculoaponeurotic anterolateral walls from the peritoneal membrane. Underneath the packing tissue is a fascial layer (endoabdominal fascia) that covers the muscles of the back and this is continuous with the transversalis fascia anterolateral on both sides.

The abdominal cavity is subdivisible into the abdomen proper and the pelvis. When viewed from one side of the sagittal plane, the pelvis lies below and behind the abdominal cavity as a basin-shaped extension with a centrally sloping muscular floor (pelvic diaphragm) part of which (the pubococcygeus) is continuous with the muscular coat of the rectum. The upper boundary of the abdominal cavity is made up of the diaphragm and, because of this, a substantial part of the abdominal cavity (almost equal to the thoracic cavity) lies under the lower rib cage. The anterolateral walls of the abdomen are made up of the musculoaponeurotic layer: the recti muscles on either side of the midline anteriorly separated by a median raphe (linea alba) and three pairs of flat wide muscles (external oblique, internal oblique and transversus abdominis).

The peritoneal cavity itself consists of the greater sac (or general peritoneal cavity) and the lesser sac (also known as the omental bursa), which lies behind the stomach, lesser omentum and transverse colon/mesocolon, and below the inferior surface of the liver. The two sacs communicate on the right side via a small slit (foramen of Winslow), the anterior fold of which, sometimes referred to as the hepatoduodenal ligament, contains the bile duct, common hepatic artery and the portal vein. It provides a readily accessible site for temporary occlusion of the hepatic inflow vessels during hepatic resections and in the event of bleeding from the hepatic vascular parenchyma or the hepatic arteries (Pringle's manoeuvre).

Anterolateral abdominal wall

The subcutaneous fat becomes divisible into two layers in the lower part of the abdomen: superficial fatty Camper's fascia and the deep membranous Scarpa's fascia that inserts into the fascia of the thigh below and parallel to the inguinal ligament. The deep fascia of the anterolateral wall is thin and covers the superficial abdominal muscles. The two musculoaponeurotic halves of the anterior abdominal wall are joined by the midline linea alba, which essentially consists of interwoven fibres of the rectus sheath forming a distinctive pattern. At the umbilical pit the skin is adherent to the linea alba and this marks the narrowest point of the abdominal wall (consisting of skin, linea alba and attachment of the ligamentum teres).

Thinning and stretching of the linea alba especially in females after multiple pregnancy results in separation (divarication) of the two rectus muscles with an intervening bulge. In the upper half of the anterior abdominal wall, the rectus muscles have both an anterior and a posterior sheath. This arrangement changes at the level of the anterior superior iliac spines where the posterior sheath is absent with its lower margin forming a crescentic outline referred to as the arcuate line (linea semicircularis). Hence, below this point, the abdominal wall is weaker. Internally the anterior abdominal wall is covered by the transversalis fascia, which is separated from the parietal pneumoperitoneum by intervening areolar fatty tissue. The fat of this layer tends to become more abundant and the connection between the peritoneal membrane and the abdominal wall looser in the lower abdomen and pelvis.

Just above the umbilicus, the parietal peritoneum is carried backwards as a two-layered membrane (falciform ligament) that is attached to the anterosuperior surface of the liver and the diaphragm. It contains a variable amount of fat (excessive and pendulant in the obese) and along its lower margin the round ligament or ligamentum teres, which is attached to the recessus of Rex at the bottom of the umbilical fissure of the liver. It represents the obliterated left umbilical vein, which connects with the left portal vein in the fetus. The muscles and skin of the anterolateral abdominal wall are supplied by the following nerves, which run in the neurovascular plane between the internal oblique and transversus abdominis muscles:

- thoracoabdominal (inferior intercostal) nerves (T7–T11)
- subcostal nerves (T12)
- ilioinguinal and iliohypogastric nerves (LI).

Below the umbilicus five structures lying between the peritoneum and the parieties form ridges or folds separating shallow fossae. In the midline, there is the median umbilical ligament (urachus, remains of the allantois), which forms a slender fibrous band between the apex of the urinary bladder and the umbilicus. The medial umbilical ligaments (obliterated portions of the umbilical arteries) run upwards and medially one on either side of the median umbilical ligament with which they merge at the umbilicus. The inferior epigastric vessels arising from the external iliac arteries run upwards and medially lateral to the medial umbilical ligaments to supply the rectus muscles and eventually enter the rectus sheath beneath the arcuate line. They form the lateral umbilical folds. Inadvertent damage to these vessels especially during laparoscopic surgery can lead to substantial haematoma formation. The named peritoneal fossae on either side are:

- supravesical: between median and medial umbilical folds
- medial inguinal: between the lateral and medial umbilical folds (site of direct inguinal hernia)
- lateral inguinal: lateral to the lateral umbilical folds (inferior epigastric arteries), includes the deep inguinal ring (site of indirect inguinal hernia).

The iliopubic tract is a condensation of the transversus abdominis fascia overlying the inner surface of the inguinal ligament when viewed from the peritoneal side (laparoscopic surgery) and separates the inferior edge of the deep inguinal ring from the femoral canal and vein below. In the early days of laparoscopic hernia surgery, this region was referred to as the triangle of doom in view of iatrogenic damage to the femoral vessels and nerve when the surgeon dissected or clipped below the iliopubic tract.

Inguinal and femoral canals

Inguinal canal

This is an oblique channel above and parallel to the medial half of the inguinal ligament and averages 4 cm in length. It contains the spermatic cord in the male and the much more tenuous round ligament of the uterus in the female – hence the inguinal canal is intrinsically weaker in the male accounting for the much higher incidence of inguinal hernia. The canal starts at the internal (deep) inguinal ring situated just lateral to the inferior epigastric artery at the midpoint of the inguinal ligament, where the transversalis fascia extends as a cone inside the internal ring and actually forms the internal spermatic fascia (innermost lining of the spermatic cord).

The vas deferens (round ligament in the female) and the testicular vessels enter the inguinal canal at the deep ring and these together with the artery to the vas deferens (derived from the inferior vesical artery), cremasteric artery (from inferior epigastric), genital branch of the genitofemoral nerve, pampiniform plexus, sympathetic fibres and lymphatic vessels form the contents of the spermatic cord. This acquires two further sheaths in its passage down the inguinal canal on its way to the scrotum – the cremasteric fascia (from the internal oblique) and the external spermatic fascia (from the external oblique aponeurosis).

The other important structure in the inguinal canal in both sexes is the ilioinguinal nerve, which lies external to the spermatic cord in the male. The external or superficial inguinal ring is a slit-like opening in the external oblique aponeurosis situated just above and lateral to the pubic tubercle. Its margins are known as the crura (medial and lateral) and are held together by means of the intercrural fibres. Normally the spermatic cord exits the inguinal canal beneath these intercrural fibres, which are supposed to prevent stretching of the superficial inguinal ring. The anterior wall of the inguinal canal is made up of the aponeurosis of the external oblique; the posterior wall by the transversalis fascia throughout reinforced medially by the fibres of the conjoint tendon.

There is considerable controversy regarding the importance of the transversalis fascia in contributing to the strength of the posterior wall of the inguinal canal, and, hence, to the pathogenesis of inguinal hernia, with some regarding it as a crucial factor that must be taken into consideration during fascial repair. Others consider that the strength of the posterior wall of the canal is largely dependent on the conjoint tendon. The arching fibres of the internal oblique and transversus abdominis form the roof of the inguinal canal, whereas the floor is made of the grooved inguinal ligament (Figure 21.1). This expands medially to form the lacunar ligament – a crescentic extension from the inguinal ligament to the pectineal line of the pubis that forms the medial margin of the femoral ring/canal.

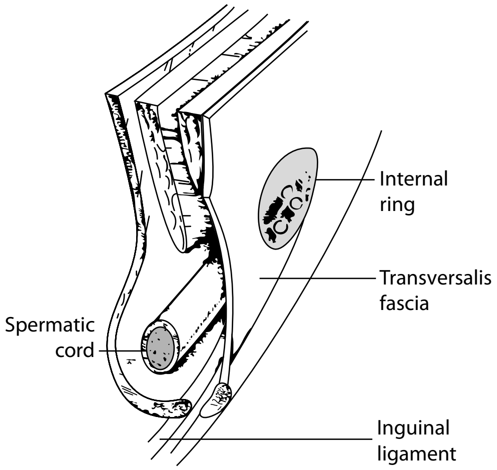


Figure 21.1 The internal view of the deep inguinal ring and cord contents. [Derived from Nyhus LM, London RE (eds). *Hernia*, 2nd edn. Philadelphia, PA: Lippincott, 1978.]

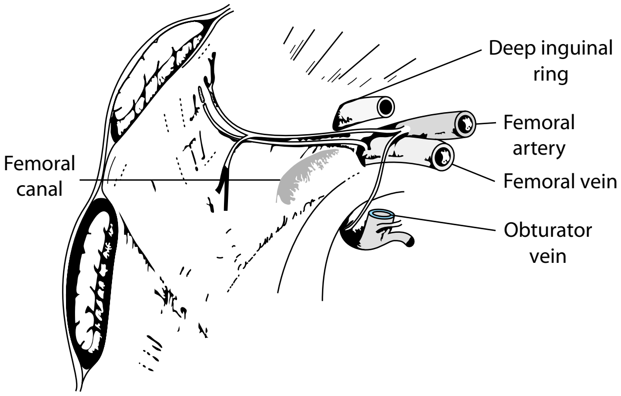


Figure 21.2 The inner aspect of the inguinal region which demonstrates the position of the femoral canal in relation to the femoral vessels and obturator foramen. [Derived from Nyhus LM, London RE (eds). *Hernia*, 2nd edn. Philadelphia, PA: Lippincott, 1978.]

Femoral canal

The cone-shaped femoral fascial sheath (derived from the endoabdominal fascia) extends into the thigh below the inguinal ligament and has three components from lateral to medial – lateral compartment (femoral artery), intermediate compartment (femoral vein) and medial compartment (femoral canal). The femoral canal itself is funnel shaped with an average length of 1.25 cm and with its oval base (the femoral ring) upwards (facing the abdominal cavity). Below it tapers to the junction of the long saphenous with femoral vein. The femoral canal normally contains loose connective tissue and a few lymphatics, and sometimes one of the deep inguinal nodes (Figure 21.2).

The boundaries of the femoral ring are the sheath covering the femoral vein laterally, the edge of the lacunar ligament medially, the inguinal ligament anteriorly and the superior pubic ramus and pectineus posteriorly. Normally it is sealed by a condensation of extraperitoneal fatty tissue known as the femoral septum, which is covered superiorly by peritoneum.

Laparoscopic anatomy of the anterior abdominal wall

Detailed knowledge and familiarity with the laparoscopic anatomy of the infraumbilical region of the anterior abdominal wall has assumed great surgical importance since the advent of laparoscopic hernia repair, which has largely replaced open repair especially for direct and indirect inguinal hernia in most centres worldwide in view of its advantages over the open surgical approach. This knowledge is essential both for ensuring good and lasting repair of the hernia without recurrence and for avoiding complications, all of which are avoidable. When viewed laparoscopically, the infraumbilical region of the anterolateral abdominal wall with the parietal peritoneal has the following landmarks consisting of folds and fossae created by the folds, which provide orientation during hernia repair:

- Folds:
  - Median umbilical ligament: ascends in the median plane to the umbilicus from the apex of the urinary bladder (represents the obliterated allantoic duct).
  - Medial umbilical ligaments, one on either side of the median umbilical ligament (represents the obliterated umbilical artery on both sides and can be traced down to the internal iliac artery).
  - Lateral umbilical ligaments, one on either side and lateral to the medial umbilical ligament consisting of a ridge of peritoneum over the inferior epigastric arteries accompanied by two veins, which skirt the medial border of the internal ring and pass upwards to the posterior rectus sheet and muscles on either side of the abdominal wall.
- Fossae (one on each side), which are formed by shallow depressions between the folds:
  - Supravesical fossa: between the median and medial umbilical folds.
  - Medial umbilical fossa: between the medial and the lateral umbilical ligaments – site of origin of direct inguinal hernia.
  - Lateral umbilical fossa: lateral to the lateral umbilical ligament – site of origin of indirect inguinal hernia (Figure 21.3).

Anatomy of the preperitoneal (extraperitoneal) space

This is essentially the anatomy of the preperitoneal space, a detailed knowledge of which is essential for the safe execution of both approaches for laparoscopic (posterior) inguinal hernia repair: TAPP (Trans-Abdominal PrePeritoneal) and TEP (Total Extra Peritoneal). The most important component of the preperitoneal space, apart from it vessels and nerves, is the transversalis fascia, which provides strength to the groin between the arch of the transversus abdominis muscle superiorly and the inguinal ligament inferiorly. The transversalis fascia consists of two layers: anterior, adherent to the rectus muscle, and posterior, which faces the parietal peritoneum (space of Borgos). The two fuse near their insertion in the pectineal (Cooper's) ligament. The anterior layer of the transversalis fascia is adherent to the rectus muscle and faces the peritoneum. The iliopubic tract and ileopectineal arch are formed by condensations of the transversalis fascia. The iliopubic tract extends from the superior pubic ramus to the anterior superior iliac spine with the inguinal ligament but distinct from it, and appears as a white line forming the superolateral boundary of the

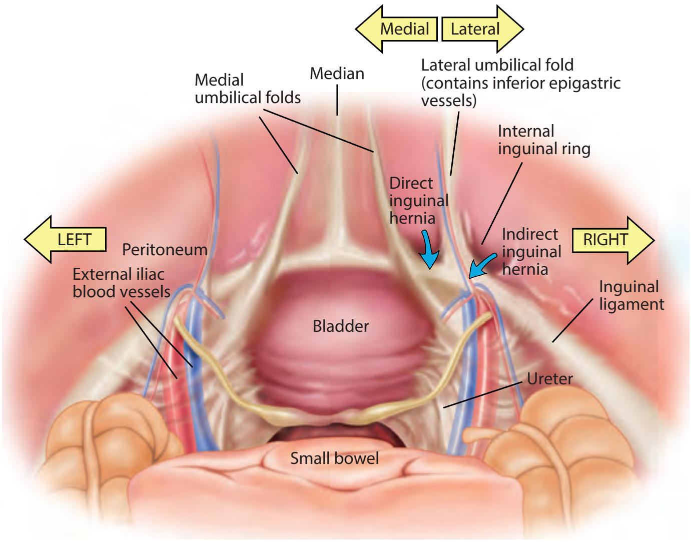


Figure 21.3 Laparoscopic view of the anterior abdominal wall.

'triangle of pain', the inferomedial border of which is formed by the gonadal vessels: the triangle of pain is so called in view of risk of nerve injury/entrapment if anchorage of the mesh (by staples) is performed within this triangle. The ileopectineal arch is a tough fibrous arch that attaches to the inguinal ligament and provides partial insertion for the internal oblique and transversus abdominis muscles and provides support for the lateral part of the groin.

Limits of the preperitoneal space

The posterior layer of the transversalis fascia separates into two compartments: (1) anterior, also known as the vascular space, and (2) posterior, known as the space of Bogros after the French anatomist who first described it. Medially the preperitoneal space is continuous with the prevesical space (space of Retzius), the dissection of which is necessary during laparoscopic hernia repair to provide sufficient space to enable good overlap of the mesh of the hernia defect.

The nerves most commonly injured during laparoscopic (posterior) inguinal hernia repair together with their manifestations are:

- Lateral femoral cutaneous nerve (commonest injury) – pain and numbness in the upper lateral aspect of the thigh known as meralgia paraesthetica.
- Femoral branch of the genitofemoral nerve – sensory changes in the femoral triangle, scrotum and labia (numbness/paraesthesia), reduced cremasteric reflux, ejaculatory dysfunction and hip pain on walking.
- Intermediate cutaneous branch of the anterior branch of the femoral nerve – pain and hyperaesthesia in the groin and thigh, weakness/atrophy of the quadriceps, painful hip joint movement. Hyperextension reproduces the pain.

These nerves are usually injured by entrapment during mesh fixation. Other less common injuries to the nerves may be caused by pressure. These include traction injuries to the ilioinguinal, iliohypogastric and the genital branch of the genitofemoral nerves.

In the performance of TAPP the surgeon must identify the following structures on entry into the peritoneal cavity:

- inferior epigastric vessels
- the two medial umbilical ligaments
- the spermatic vessels
- the vas deferens
- trapezoid of disaster: area at the meeting of the vas deferens and testicular vessels where the external iliac vessels can be injured.

The inferior epigastric vessels enable the differentiation of direct from indirect inguinal hernia. The importance of identification of the medial umbilical ligament is for the surgeon to avoid extending the dissection medial to it as this may risk injury to the urinary bladder. The spermatic vessels form the medial border of the triangle of doom and the lateral border of the triangle of pain. The vas deferens forms the medial border of the triangle of doom/trapezoid of disaster.

Following opening of the peritoneal trapdoor in TAPP, the following structures have to be identified:

- the internal inguinal (spermatic) ring
- iliopubic tract
- pectineal (Cooper's) ligament
- femoral canal.

The iliopubic tract separates the inguinal from the femoral canal, which lies posterior to it. During mesh fixation, staples or sutures should not be place below the ileopubic tract.

Posterior abdominal wall and retroperitoneum

From both anatomical and functional aspects, this is best considered from its ventral aspect. In the midline are the bodies, transverse processes and the intervertebral discs of the five lumbar vertebrae. Laterally the musculoaponeurotic wall (external and internal oblique and transversus abdominis muscles) extends from the twelfth rib to the pelvic brim. The psoas muscles take origin from the bodies and transverse processes of the upper four lumbar vertebrae and are joined by the iliacus muscles on their lateral aspect. Posteriorly and laterally, the quadratus lumborum muscles form the remaining support for the lumbar nerves and vessels (Figure 21.4). The lower part of the posterior abdominal wall is made up of the iliacus muscles covering the iliac bones of the bony pelvis. The complex anatomy of the posterior abdominal wall is of great importance during surgery on retroperitoneal organs and retroperitoneal tumours. In this respect, the muscles of the posterior abdominal wall are covered by the endoabdominal fascia, which is given various names depending on the underlying muscles:

- Psoas fascia or sheath covering the psoas major muscles – attached medially to the lumbar vertebrae and below to the pelvic brim.
- Quadratus lumborum fascia – attached to the transverse processes of the lumbar vertebrae and continuous laterally with the anterior layer of the thoracolumbar fascia.
- Thoracolumbar fascia – splits to enclose the deep muscles of the back and is thick and strong in the lumbar region, where it extends from the costal margin to iliac crests and is attached laterally to the internal oblique and transversus abdominis muscles (Figure 21.5).

The retroperitoneal space is bounded above by the diaphragm and below by the pelvic brim but is of course continuous with the retroperitoneal space of the pelvis. The space contains the aorta, vena cava, cisterna chyli, para-aortic

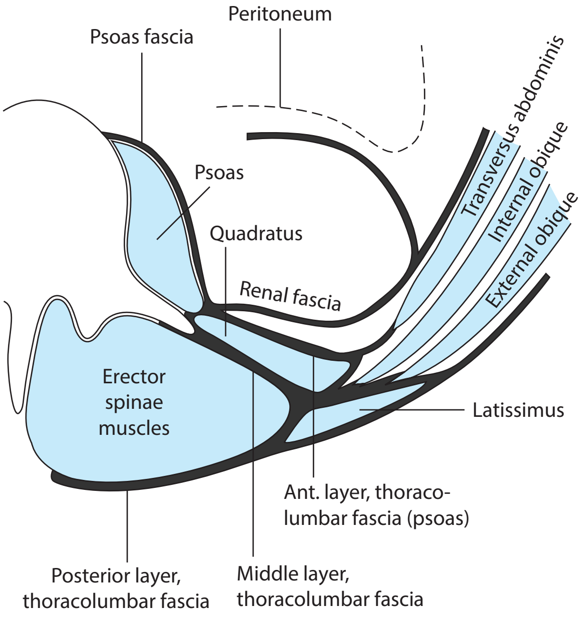


Figure 21.5 Thoracolumbar fascia. This splits to enclose the deep muscles of the back and is thick and strong in the lumbar region, where it extends from the costal margin to iliac crests and is attached laterally to the internal oblique and transversus abdominis muscles.

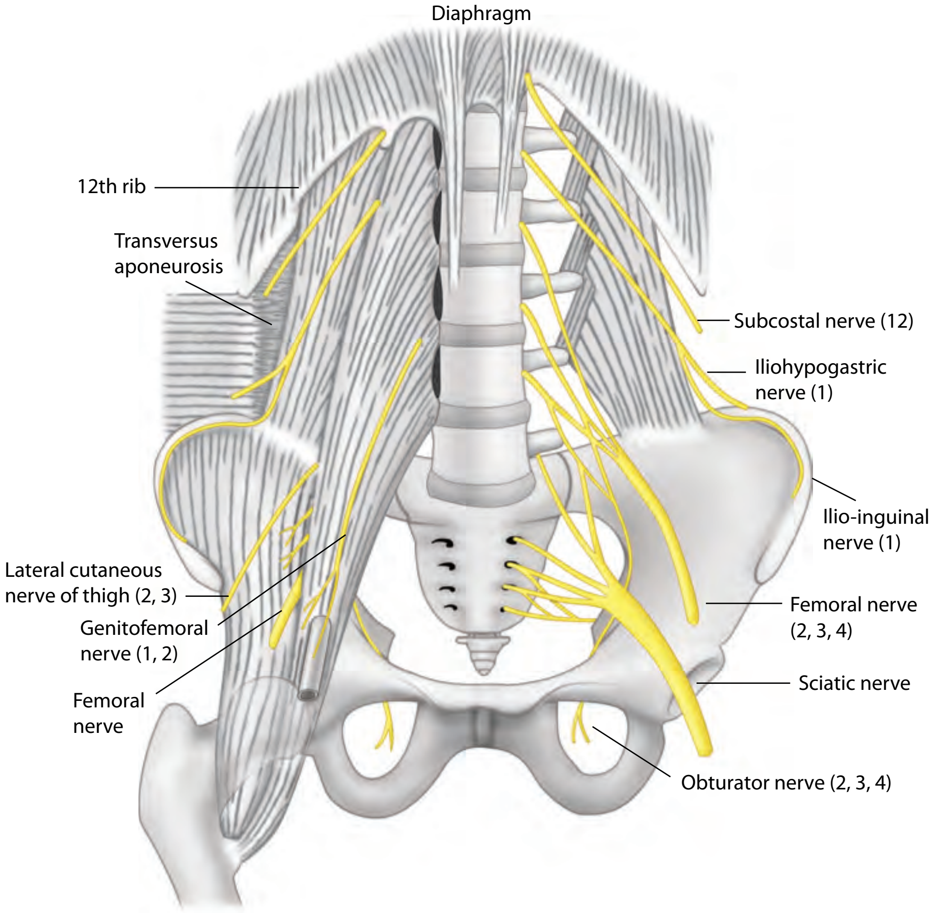


Figure 21.4 The anatomical features of the posterior abdominal wall.

glands and vessels, the kidneys and ureters, the adrenal glands, the second and third parts of the duodenum, the lower end of the bile duct, pancreas, the various nerve plexuses and the lumbar sympathetic chain.

The somatic nerves of the posterior abdominal wall are the:

- subcostal nerves
- lumbar nerves
- lumbar plexus – located in the posterior part of the psoas major (L1–4).

In turn, the lumbar plexus gives rise to the following nerves (on either side):

- *Obturator* nerve (L2–4): exits from the medial border of the psoas and descends medially through the pelvis to reach the thigh, where it supplies the adductor muscles.
- *Femoral* nerve (L2–4): exits from the lateral border of the psoas major, supplies the iliacus and then exits the abdomen deep to the inguinal ligament lateral to the femoral artery. Supplies hip flexor muscles and knee extensors.
- *Lumbosacral trunk*: crosses the sacral wing to enter the pelvis to form the sacral plexus together with the S1–4 nerves.
- *Genitofemoral* nerve (L1, L2): exits from the anterior surface of the psoas major but stays deep to the psoas fascia and lower down (lateral to the common iliac arteries) it divides into its femoral and genital branches.
- *Ilioinguinal and iliohypogastric* nerves (both from L1): descend in front of the quadratus lumborum, penetrate the transversus abdominis near the anterior superior iliac spine and then travel through the internal and external oblique muscles (to which they give branches) to the skin of the groin and pubic region. Only the ilioinguinal nerve courses through the inguinal canal on its way to the skin.
- *Lateral femoral cutaneous* nerve (L2, L3): runs downwards and laterally across the iliacus muscle and exits the abdomen below the inguinal ligament and close to the anterior superior iliac spine. It supplies the skin of the anterolateral aspect of the thigh.

The autonomic nerves consist of:

- vagal branches
- abdominopelvic splanchnic nerves – thoracic (T5–12) and lumbar (L1–3)
- prevertebral sympathetic ganglia
- abdominal autonomic plexuses.

The lumbar splanchnic nerves (three or four in number) enter three nerve plexuses (intermesenteric, inferior mesenteric and superior hypogastric), which supply presynaptic fibres to the prevertebral ganglia and plexuses, situated close to the major branches of the aorta (coeliac, superior mesenteric, aortorenal and inferior mesenteric). The postsynaptic fibres from these prevertebral ganglia form periarterial plexuses that supply the various organs.

The parasympathetic supply to the various plexuses is derived from the vagi and the pelvic splanchnic nerves, which originate directly from the ventral rami of S2–4. The location of the abdominal nerve plexuses is as follows:

- *Coeliac* (solar): surrounds the coeliac trunk.
- *Superior mesenteric*: around the origin of the superior mesenteric artery.
- *Inferior mesenteric*: surrounds the corresponding artery as it arises from the lower aorta.
- *Intermesenteric:* on the anterior surface of the aorta between the superior and inferior mesenteric arteries.
- *Superior hypogastric:* continuous with the intermesenteric and inferior mesenteric plexi and located over the bifurcation of the aorta. It gives rise to the *hypogastric nerves/plexi,* which continue towards the pelvic floor lateral to the rectum, urinary bladder and uterine cervix. These nerves/plexuses must be preserved in males during rectal and prostatic pelvic surgery because they are responsible through their parasympathetic fibres (S2–4) for erection and ejaculation.

Trauma to the space from blunt or penetrating injury is common, particularly in high-speed vehicle accidents. Damage to the contained structures leads to haemorrhage and haematoma formation (see Chapter 14). Haemorrhage into the retroperitoneal space may also occur spontaneously from a leaking abdominal aneurysm, and in patients with bleeding disorders and those taking anticoagulants.

Pelvic floor

This is largely made of the levator ani muscle, which slopes downwards from the lateral pelvic walls, pubic bones, ischial spines and coccygeus towards the centre of the pelvis. Together with the coccygeus muscles posteriorly, the levator ani muscle forms the pelvic diaphragm through the centre of which pass the rectum and urethra in both sexes, separated by the vagina in the female. The levator ani, which has a separate nerve supply (S3, S4), has three components: the pubococcygeus (stretching from the pubis to the coccyx), the puborectalis (which forms a sling around the anorectal junction) and the ileococcygeus (which forms the posterior part). Aside from forming the floor of the pelvis and providing support for the rectum, bladder and uterus, the levator ani plays an important part in the voluntary control of micturition and defaecation.

The pelvic diaphragm is covered by fascia on both its upper and lower surfaces. The ischiorectal fossa is the space between the sloping inferior surface of the pelvic diaphragm and the ischial tuberosity. It is normally packed by fatty tissue and is traversed by the inferior rectal vessels (Figure 21.6). The important nerves associated with the pelvic floor are the pelvic splanchnic (S2–4), nerves to the levator

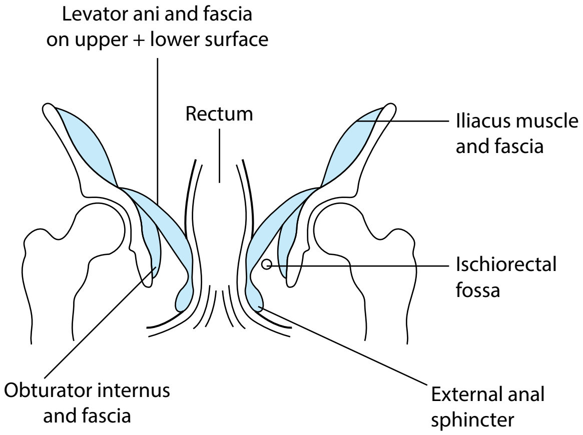

Levator ani and fascia on upper + lower surface
Rectum
Iliacus muscle and fascia
Ischiorectal fossa
Obturator internus and fascia
External anal sphincter

Figure 21.6 The pelvic diaphragm is covered by fascia on both its upper and lower surfaces. The ischiorectal fossa is the space between the sloping inferior surface of the pelvic diaphragm and the ischial tuberosity. It is normally packed by fatty tissue and is traversed by the inferior rectal vessels.

ani and coccygeus muscles and the pudendal nerve. The pudendal nerve (S2–4) exits the pelvis through the greater sciatic foramen to enter the pudendal canal on the medial side of the ischial tuberosity. Although it is the chief sensory nerve to the external genitalia in both sexes, it supplies muscular branches to the external anal and urethral sphincters and to the perineal muscles.

The umbilicus

The umbilicus normally lies in the plane of the disc between the third and fourth lumbar vertebrae, although clearly this varies with sex, age and degree of obesity. When examined from the deep surface, four fibrous cords are seen radiating from it, and these represent structures which in fetal life traverse the umbilical cord: the umbilical vein, right and left umbilical arteries and the urachus. The additional structure to pass through the cord in embryonic development is the vitellointestinal duct with the accompanying vitelline veins. Not surprisingly, subclinical defects in the umbilical cicatrix are common and these give rise to paraumbilical hernias, usually in obese females.

Omphalitis

Infection of the umbilicus is fortunately a rare complication of the neonate. Faulty technique in dealing with the umbilical cord at birth leads to infection by *Staphylococcus aureus* or haemolytic streptococcal organisms with local suppuration and cellulitis of the adjacent abdominal wall. Treatment with appropriate antibiotics and local drainage for bacteraemia is common. There is an association between umbilical sepsis and portal vein thrombosis with the development of portal hypertension in these children.

In the adult, infection of the umbilicus is a more chronic condition with a seropurulent, often foul smelling, discharge and the development of granulation tissue. In the majority of patients, the fault lies in a lack of hygiene, but occasionally small foreign bodies are present and these include an ‘umbilical stone’, a concretion due to desquamated dead superficial epidermal cells. Pilonidal cysts can also occur, albeit rarely in the umbilical pit.

Congenital abnormalities

Patent urachal remnant

A persistent discharge of urine from the umbilicus results from a completely patent urachus. This complication may be delayed into childhood or early adult life, but, in these circumstances, there is almost always an underlying urinary obstruction that has forced open the near-obliterated urachal remnant. It is important that prior to excision of the urachal remnant, which can be preformed laparoscopically, the bladder should be shown to be free from any form of obstruction, e.g. posterior urethral valves.

Two other urachal abnormalities may be delayed into adult life: urachal sinus and urachal cysts. In the former, an intermittent umbilical discharge is the prominent symptom; in the latter, the development of a very tender infected swelling in an infraumbilical position leads to exploration and demonstration of an infected urachal cyst. In both instances excision is indicated but repair of the bladder is not usually required.

Vitellointestinal duct remnant

The persistence of this remnant may show as a faecal fistula at the umbilicus, a bud or polyp of intestinal mucosa at the umbilicus, a cyst lying deep to the umbilicus between the abdominal wall and the ileum, a deep umbilical sinus, or a Meckel’s diverticulum with or without a fibrous band attached to the posterior aspect of the umbilicus and responsible for intestinal obstruction by volvulus or band obstruction (Figure 21.7).

The umbilical sinus, polyp or cyst require excision. The completely patent vitellointestinal duct may be complicated by partial or complete prolapse of the ileum through the umbilicus and requires early correction before this complication occurs. Meckel’s diverticulum is discussed elsewhere (see Chapter 29).

Abdominal wall defects

A failure of the umbilical defect to close leads to the development of omphalocele or an umbilical hernia. Major defects in the development of the abdominal wall are associated with exstrophy of the bladder and other forms of developmental abnormality including intestinal malrotation. The prune belly syndrome (PBS) is a rare genetic disorder affecting about one in 40 000 births, predominantly male infants (97%). PBS is a disorder of the urinary system, characterized by wrinkled abdominal wall skin associated with congenital absence/hypoplasia of abdominal muscles (hence also known as abdominal muscle deficiency syndrome and various other names), major urological abnormalities and cryptorchidism. The abdominal contents are covered only with skin, peritoneum and an intervening hypoplastic muscular layer. The urological abnormalities include a long dilated prostatic urethra associated with prostatic hypoplasia; a large, vertically oriented thick-walled bladder; a urachal remnant from the dome of the bladder; tortuous and dilated ureters; varying amounts of hydronephrosis; and renal dysplasia.

Rectus sheath haematoma

Rectus sheath haematoma (RSH) is rare, although its incidence appears to be rising. It is caused by rupture of the superior or inferior epigastric arteries or their branches. RSH is either sporadic, when there is usually a predisposing cause, or traumatic insult (blunt and stab injuries) and following abdominal surgery (iatrogenic). There has been a significant increase in the incidence of iatrogenic RSH with the advent of laparoscopic surgery owing to failure to take note of the surface marking of the superior and inferior epigastric arteries and veins during blind placement of abdominal ports. This iatrogenic complication is entirely preventable.

Clinical features

RSH is two to three times commoner in females. The median age at presentation is 45 years and the incidence rises with advancing age. The commonest predisposing cause in sporadic cases is anticoagulation therapy. The haematoma is usually located behind the rectus muscle as both arteries enter the muscular compartment posteriorly. Sporadic RSH usually arises as a result of vigorous contractions of the rectus

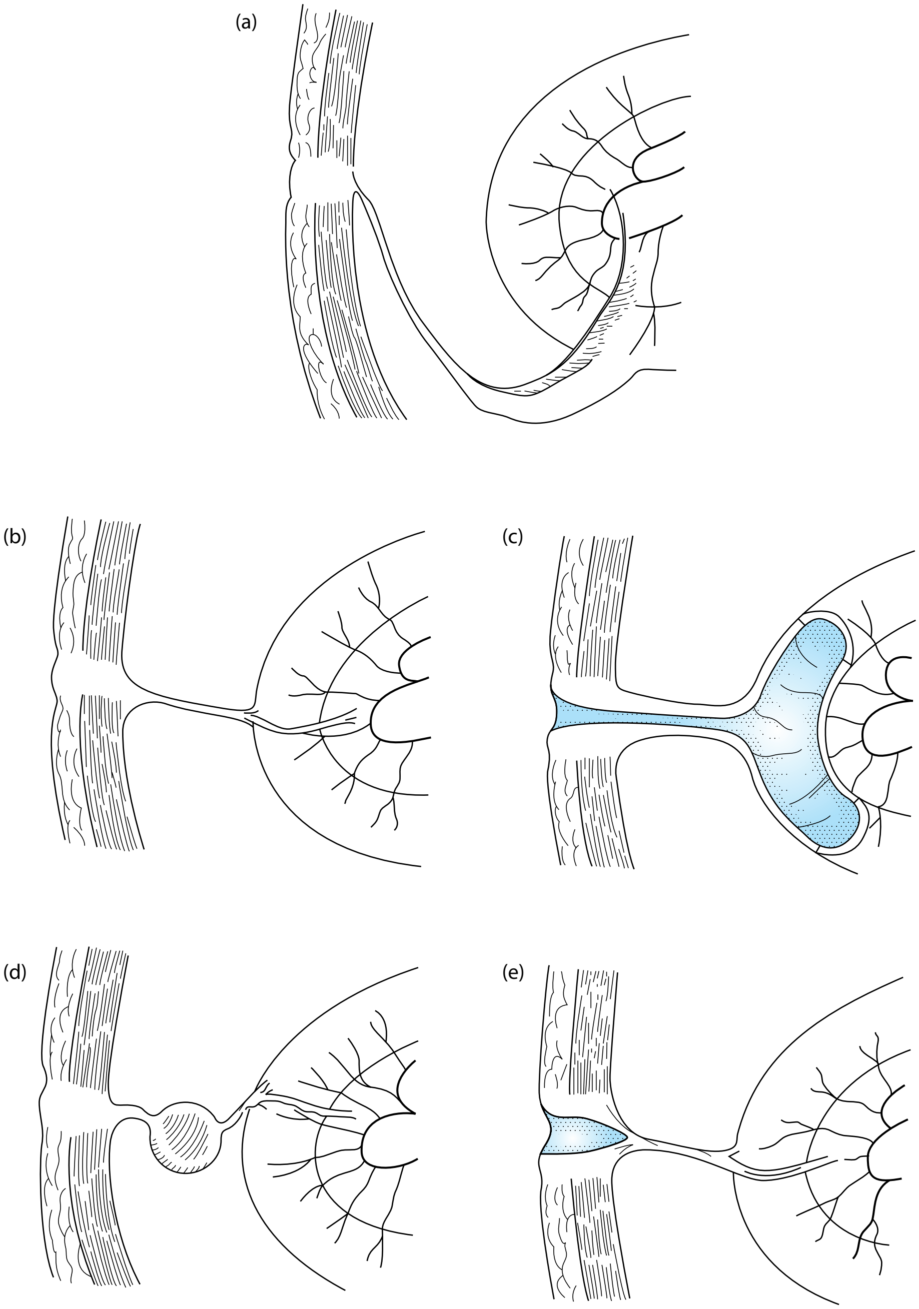


Figure 21.7 The vitellointestinal duct may connect with the umbilicus as: (a) a fibrous cord extending from a Meckel's diverticulum; (b) a simple fibrous cord to a loop of ileum; (c) an umbilical intestinal fistula; (d) a fibrous cord with cyst; (e) an umbilical sinus.

muscles from strenuous vigorous exercise, severe coughing, vomiting and straining at stool. The history in the sporadic variety is usually that of strenuous exercise, strain or paroxysm of coughing followed by severe abdominal pain accompanied with nausea, anorexia and vomiting. On examination an acutely tender rectus muscle swelling can be felt in the abdominal wall. A mild pyrexia and leucocytosis may occur. Sixty per cent of RSHs are situated on the right side and >80% in the lower quadrants. Haematomas above the arcuate line are caused by damage to the superior epigastric artery or its perforating branches, whereas the inferior epigastric vessels are involved in cases below the arcuate line. Haematomas below the arcuate tend to be more extensive and bleed more as the rectus abdominis muscle is only supported posteriorly by the transversalis fascia and the parietal peritoneum. These RSHs often cross the midline and become bilobar. RSH near the peritoneum can result in peritoneal irritation with abdominal rigidity and gastrointestinal symptoms. Classically, three eponymous signs are described for the diagnosis of RSH:

1 Fothergill sign: determines whether an abdominal mass is within the abdominal wall or intra-abdominal: voluntary contraction of the recti muscles fixes the mass and makes it

more painful/tender. The test is difficult to elicit conclusively in obese or pregnant females.

2 Carnett sign: also differentiates between abdominal wall as distinct from intra-abdominal pathology. Pain or tenderness is elicited when the supine patient lifts his or her head or shoulders from bed.

3 Cullen sign: periumbilical ecchymosis associated with retroperitoneal or abdominal wall haemorrhage – in RCH bruising appears after 2–5 days and often extends into the flanks.

The mass is usually palpable, painful, firm, non-pulsatile and does not move with respiration. In obese patients, the mass may be difficult to palpate because of the posterior location of the haematoma behind the rectus. Signs of local peritoneal irritation with rebound tenderness and guarding may be present, especially with large RSHs which cross the midline situated in the subumbilical region. Rarely, a haematoma may rupture into the peritoneal cavity causing haemoperitoneum.

The diagnosis of RSH is confirmed by an ultrasound scan of the abdominal wall (Figure 21.8). Ultrasonography is used as a first-line diagnostic test for RSH and also to monitor the condition in cases treated conservatively.

However, CT or MRI provide better definition of the pathology and especially on the need for conservative or active surgical treatment in sporadic patients. Three types of RSH are identified by CT in terms of severity of the disease:

- *Type I.* Unilateral intramuscular haematoma with increase in size of the muscle but without any dissection of fascial planes. Patients present with mild-to-moderate abdominal pain. Type I haematomas resolve spontaneously.
- *Type II.* Intramuscular, often bilateral, haematoma but with blood extravasation between muscle and the transversalis fascia. May be accompanied by a fall in haematocrit: the patient requires close observation and may need blood transfusions but the condition is treated conservatively in the first instance.
- *Type III.* Large haematoma which extends outside the muscle to the transversalis fascia and transversus muscle, the peritoneum and the prevesical space. May be accompanied by haemoperitoneum: the patient is often hypovolaemic and requires blood transfusion. Reversal of anticoagulation is needed for patients on anticoagulation; patients require urgent surgical treatment, especially if they are unstable.

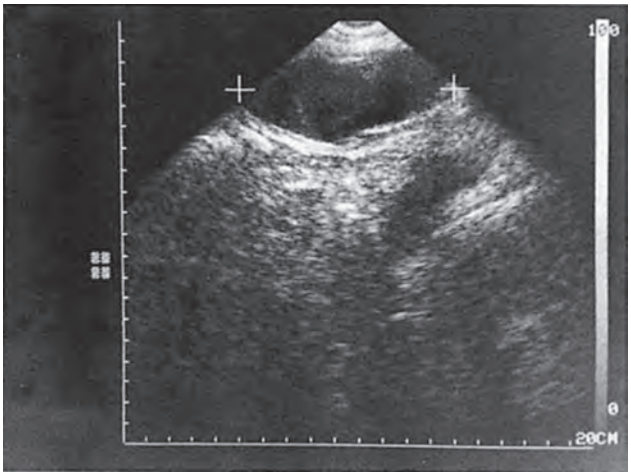


Figure 21.8 Ultrasound of the lower abdominal wall of a patient on dicoumarin therapy, who developed severe abdominal pain after a coughing spasm. The area next to the '+' shows a haematoma confined by rectus sheath.

Conservative and non-surgical treatment

In some centres CT staging is used for triage with sporadic type I being treated conservatively with analgesia and bed rest as an outpatient. Types II and III require hospital admission for close observation and possible active treatment by surgery or radiological intervention. The serious cases are those in the CT type III category on anticoagulation therapy and account for most of the reported fatal cases. Anticoagulation therapy should be stopped immediately on diagnosis and reversal to normal clotting and prothrombin time (PT) achieved. Patients on heparin are reversed with protamine sulphate, whereas patients on oral anticoagulants are commenced on vitamin K analogues but may need fresh frozen plasma if they continue to bleed as the return to a normal PT may take several days. The patients are monitored closely for evidence of hypovolaemia, enlargement of RSH and signs of peritonism.

Radiological intervention with arterial transcatheter Gelfoam/coil embolization is used in some centres to achieve haemostasis and decrease the need for blood transfusion. It can also be used as an alternative to surgery in frail patients not responding to conservative management.

Surgical treatment

Surgical treatment is advisable for failed conservative treatment of sporadic RSH and in all cases caused by trauma or surgery. In sporadic cases the incision is placed on the summit of the haematoma. Surgical treatment includes complete evacuation of the clot, ligation of bleeding vessels and repair of the rectus sheath. Most surgeons opt not to insert drains if the haemostasis is complete. In patients with stab wounds, laparoscopy is the quickest way to determine breach of the parietal peritoneum and intra-abdominal bleeding, which would require full exploration. If the peritoneum is intact, the haematoma is treated as in the sporadic cases. Cases following blunt trauma require CT assessment before surgery to exclude associated intra-abdominal injuries.

Disorders of the peritoneum

Primary peritonitis

In surgical practice, peritonitis is usually *secondary* to gastrointestinal perforation/injury/anastomotic dehiscence or a gangrenous/infected hollow visceral organ. Bacterial peritonitis can, however, be primary, i.e. develops in the absence of surgery/trauma or any primary intra-abdominal focus of infection. This is usually encountered in patients with chronic disease in whom the term *spontaneous bacterial peritonitis* (SPB) is used as opposed to the rarer *primary bacterial peritonitis*, which is reserved for the disease occurring in previously normal individuals (usually women and children).

A fourth category of peritonitis is recognized – *tertiary peritonitis* (TP). This occurs in intensive care patients and is defined as the persistence or recurrence of intra-abdominal infection following apparently adequate therapy of primary or secondary peritonitis. In one reported series tertiary peritonitis developed in 74% of patients admitted with intra-abdominal infection to a surgical intensive care unit (ICU). Patients who develop TP have a significantly longer ICU stay and more advanced organ dysfunction, reflected in higher ICU mortality (64% versus 33%), than patients with uncomplicated secondary peritonitis. The most common infecting organisms in patients with TP are *Enterococcus, Candida, Staphylococcus epidermidis* and *Enterobacter.* The infectious foci are rarely amenable to percutaneous drainage and these patients do not benefit from laparotomy as the infection is diffuse and poorly localized. In common with nosocomial pneumonia in critically ill patients, TP appears to be more a reflection than a cause of an adverse outcome.

Primary bacterial peritonitis

By definition, primary bacterial peritonitis occurs in healthy individuals and the infecting organism is of the Gram-positive type, most commonly *Streptococcus pneumoniae* and group A streptococci. The disease is encountered in children, adolescents and adult women, in whom it may follow childbirth, chest and urinary tract infection. In most instances, the infection is haematogenous.

Infants and children who develop primary pneumococcal peritonitis usually present with acute abdominal pain, vomiting and fever, and abdominal signs indicative of peritoneal inflammation. Diagnosis is made on clinical grounds helped by abdominal radiographs and ultrasound examination. The blood cultures may be positive in these patients. Treatment is with intravenous antibiotics in the first instance. Awareness of the condition is important as primary peritonitis is a rare condition in children, and thus it will be overlooked unless it is considered in the differential diagnosis of children presenting with an acute abdomen.

The disease in adults is confined to women, and, although the infection is commonly pneumococcal, instances of gonococcal peritonitis have been reported. The typical patients are usually young adolescent girls or women of childbearing age. Some cases are reported in association with acute (non-perforated) appendicitis. Primary peritonitis caused by *S. pneumoniae* may follow childbirth. The patients become pyrexial and develop abdominal pain, diarrhoea and clinical signs of peritonitis. In addition to antibiotic therapy, laparotomy is usually necessary to remove pus and for abdominal lavage. Culture of vaginal swabs is usually positive for pneumococcus in patients who develop the condition after childbirth. The prognosis of primary peritonitis with early diagnosis and treatment (antibiotics and abdominal lavage) is good with recovery of the vast majority of patients.

Management

In many of these patients, the differentiation between primary and secondary bacterial peritonitis is difficult, if not impossible. In practice all are explored surgically and the condition is diagnosed because of the absence of a primary focus, although the surgeon must be aware of the mild (non-perforated) appendicitis associated with a seropurulent peritonitis usually in young females. These are now considered as instances of 'primary peritonitis'. At operation, adequate samples of the peritoneal fluid are taken for culture, the appendix is removed if it appears mildly inflamed and abdominal lavage is undertaken before closure. As the infection is most commonly due to *S. pneumoniae,* the initial antibiotic should consist of amoxicillin/clavulanic acid.

Spontaneous bacterial peritonitis

This carries a worse prognosis and has a definite mortality from septic shock and multiorgan system failure. It is an infection of intraperitoneal fluid (ascites or peritoneal dialysate). In general patients with a low ascitic protein concentration ($<1.0\,g/dL$) irrespective of the nature of the underlying disease (liver or renal) are prone to SBP. The groups of patients who are prone to develop SBP include:

- cirrhotic patients with ascites, Wilson disease, chronic active hepatitis
- renal failure in patients on chronic peritoneal dialysis
- patients with the nephrotic syndrome
- immunocompromised patients.

The common factor in these patients is reduced resistance to bacterial infection. The infecting organisms are often Gram-negative.

Clinical features of spontaneous bacterial peritonitis due to liver and renal disease

About 30% of patients with SBP have no symptoms or signs directly referable to the abdomen, and, therefore, a high index of suspicion must be kept especially in susceptible groups. Early diagnosis is imperative, as otherwise the mortality is high, ranging from 50% to 80%. In other patients, the disease develops insidiously and localizing signs of peritonitis are present but often minimal. The most common manifestations include abdominal pain, fever, rebound tenderness and diminished or absent bowel sounds. The full-blown picture is accompanied by septic shock and is invariably fatal. Once suspected, a 100 mL specimen of ascitic fluid is taken for culture and Gram staining of the deposit after centrifugation. The fluid is also examined for polymorphonuclear count (PMN) and pH. A PMN count >250 mL and a pH <7.37 are diagnostic and indicate the need for antibiotic therapy, even if the culture of the ascitic fluid is negative. A blood culture should also be taken and is positive in 70% of cases.

Spontaneous bacterial peritonitis in cirrhotic patients with ascites

In this instance there is bacterial infection of ascitic fluid in the absence of any intra-abdominal, surgically treatable source of infection. The reported incidence of SBP in cirrhotics with ascites varies from 18% to 25%. Most of the infections are aerobic and 50–60% of the reported cases have been caused by *Escherichia coli*. The aetiology is thought to involve:

- bacterial translocation from the gut to mesenteric lymph nodes
- depressed activity of the reticuloendothelial phagocytic system
- decreased antimicrobial capacity of ascitic fluid – low levels of C3, opsonins and fibronectin.

Diagnosis is based on clinical suspicion and analysis of ascitic fluid (white cell count and culture in blood culture bottles). Treatment is with a third-generation cephalosporin.

This achieves a cure rate >80%. Cytological cure is obtained in 65% of patients who are culture positive and sensitive to ceftriaxone after 48 hours of treatment and 95% are cured of their infection after 5 days of treatment. Even so many patients (up to 30%) die during hospitalization despite documented cure of their SBP from complications related to their end-stage liver disease (renal failure, gastrointestinal bleed, cerebral oedema). Infection-related mortality does occur, and is related to the onset of bacteraemia (often from *Pseudomonas* spp.).

Prophylactic selective intestinal decontamination with oral norfloxacin is recommended for the prevention of SBP in cirrhotic patients who are at high risk for developing SBP. These include:

- hospitalized cirrhotic patients with gastrointestinal haemorrhage
- patients with low ascitic fluid total protein (<1 g/dL)
- patients with a high bilirubin level and/or low platelet count.

The long-term prognosis of SBP patients is poor, and these patients should be considered for liver transplantation.

Spontaneous bacterial peritonitis in patients on continuous ambulatory peritoneal dialysis

As in cirrhotic patients, SPB in patients on continuous ambulatory peritoneal dialysis (CAPD) can be culture positive or negative. The infection may be caused by either Gram-positive cocci or Gram-negative bacilli. SPB in this patient group can also arise as a consequence of catheter-related infections (subcutaneous tunnel or catheter exit site). The risk factors for catheter exit-site infection documented in a prospective randomized trial by univariate analysis and multiple logistic regression analysis are:

- younger age (<50 years)
- low serum albumin level (<35 g/L)
- number of previously placed peritoneal dialysis catheters
- short cuff–exit distance (<2 cm)
- *S. aureus* nasal carriage.

The standard primary treatment of CAPD SBP is with intraperitoneal netilmicin combined with intermittent intraperitoneal vancomycin.

However, in a larger randomized clinical trial, oral levofloxacin in combination with intermittent intraperitoneal vancomycin was found to be equally effective. This regimen is simpler to administer and less costly. It is currently recommended as the primary therapy in centres with relatively low exposure and, thus, low background resistance to fluoroquinolones.

The primary cure rate of CAPD SBP averages 75% but varies considerably from centre to centre. Higher cure rates are obtained in patients with culture-negative and Gram-positive infections (75–80%) than in those with Gram-negative infections (55%).

Spontaneous bacterial peritonitis in patients with the nephrotic syndrome

SBP associated with the nephrotic syndrome is largely encountered in children, and is rare in adults. The patients usually have active nephrosis and present with diffuse abdominal pain, ascites, fever and rigors. The infection is caused by either Gram-positive or -negative pathogens, but culture-negative cases have been reported. When it is severe and unresponsive to antibiotic therapy, the patients die of septic shock.

Granulomatous peritonitis

The formation of multiple peritoneal granulomas with the development of ascites may rarely occur as a manifestation of sarcoidosis when differentiation from tuberculous peritonitis can be difficult and is based on a positive Kveim test, negative tuberculosis cultures and a lack of response to antituberculous therapy.

Starch peritonitis (starch granuloma syndrome) used to be a more common cause of granulomatous peritonitis in surgical practice and caused substantial morbidity. Talc (magnesium silicate) was the initial lubricant for surgical gloves and its implantation during surgery caused severe granuloma formation, chemical peritonitis and adhesion formation. It was replaced by Bio-Sorb in 1949. This is the epichlorohydrin polymer of cornstarch mixed with 2% magnesium oxide and small amounts of sodium sulphate and sodium chloride. When introduced, it was claimed to be completely absorbed by the peritoneal membrane and was thus free of the disadvantages encountered with talc. Subsequent experience has shown that reactions to Bio-Sorb do occur and include a syndrome of starch peritonitis (starch granuloma syndrome), which has a characteristic and well-recognized clinical picture. The disease starts 2–6 weeks after abdominal surgery with a low-grade fever, anorexia, nausea, vomiting, abdominal distension, cramp-like pain and tenderness. The abdominal distension is due to ileus and to the accumulation of ascitic fluid. Multiple granulomas develop on both the visceral and parietal pneumoperitoneum. The ascitic fluid is usually amber but may be serosanguinous and contains many leucocytes made up largely of lymphocytes and monocytes. The granulomatous nodules consist of collections of lymphocytes, macrophages, polymorphs and eosinophils around starch granules, which have a characteristic Maltese-cross appearance on microscopy (Figure 21.9). There is debate concerning the nature of the reaction. Some ascribe its development to a state of hypersensitivity to cornstarch, which can be demonstrated by skin patch testing in patients who develop the condition. Others argue that it represents a foreign body reaction. This is unlikely in view of the rare occurrence. The diagnosis is made on the clinical picture together with the demonstration of starch granules in the ascitic fluid obtained by an abdominal tap (iodine staining and polarized light microscopy). Surgery is avoided if the diagnosis is certain and treatment is conservative. Rapid resolution is seen with systemic corticosteroid therapy. The complications of starch peritonitis include sinus and fistula formation, adhesion formation and intestinal obstruction.

Bio-Sorb is seldom used in gloves nowadays, but, if it is, the following measure is advisable. After the gloves have been put on by the surgeon, 10 mL of povidone iodine (Betadine) is applied and smeared on the surface of the gloves. The black starch-iodine granules are then washed off by pouring 500 mL of sterile water from a container. In most instances, hydrogel polymers are used nowadays as lubricants for surgical gloves.

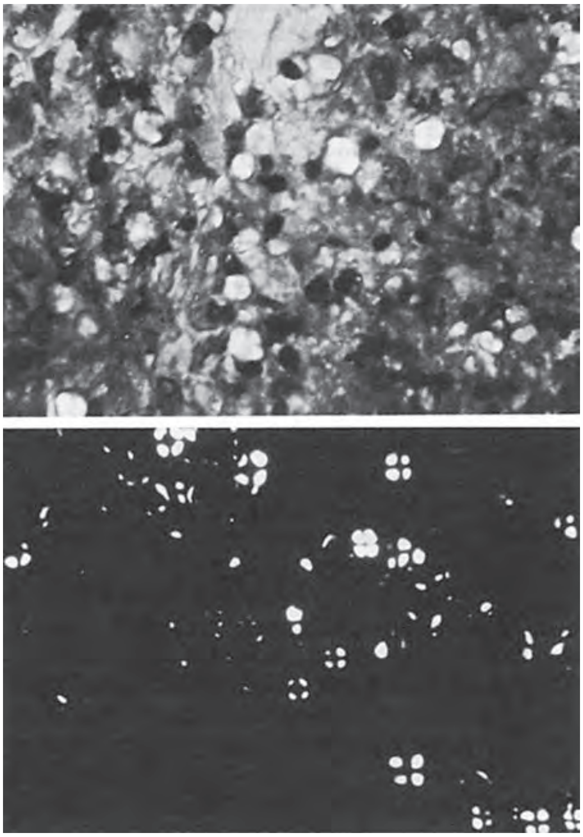


Figure 21.9 Maltese-cross appearance of cornstarch granules visualized with polarized light. (Courtesy of I. Capperauld, Ethicon Laboratories, Edinburgh, UK.)

Meconium peritonitis

Meconium peritonitis is an unusual and not uncommonly fatal form of neonatal peritonitis. It is caused by antenatal extravasation of meconium into the peritoneal cavity and this results in a serious illness characterized by intraperitoneal calcification, dense inflammatory fibrosis with or without giant pseudocyst formation. In the majority of cases, no obvious cause can be found for the meconium peritonitis. In neonates and infants in whom a cause is found at laparotomy, this is either jejunal or ileal atresia and less commonly perforation of the appendix. The most striking and common findings during surgery are gross inflammatory adhesion bands with matted intestinal loops. Giant pseudocysts or intestinal perforation are present in 50–60% of cases. The surgical treatment depends on the exact operative findings in the individual neonate/infant. Some need adhesiolysis alone. Pseudocysts are partially resected and temporary enterostomy should be done for perforated bowel or when resection is needed with no attempt at primary anastomosis.

Non-surgical pneumoperitoneum

The introduction of as little as 10 mL of air into the peritoneal cavity may be demonstrated by erect films of the abdomen and chest in the right subdiaphragmatic region. Intraperitoneal air introduced during laparotomy is rapidly absorbed in infants (within 12 hours), often takes longer in older children and adults (38–48 hours) but may persist for 3–4 weeks in adults. Studies on operating room air have indicated the presence of deleterious substances including endotoxin. The absorption of this through the serosal surfaces of the exposed bowel and peritoneal membrane has been incriminated, in part, for the postoperative 'stress response' after open surgery.

In surgical practice, pneumoperitoneum detected by an erect abdominal or chest radiograph usually indicates perforation of a hollow viscus and is accompanied by clinical signs of peritoneal irritation indicative of serious intra-abdominal disease necessitating urgent surgical intervention. However, 10% of all cases of pneumoperitoneum are not accompanied by physical abdominal signs or evidence of significant underlying disease. The common causes of this non-surgical pneumoperitoneum are:

- Escape of air from the tracheobronchial tree in patients with chronic obstructive airways disease and in patients on intermittent positive-pressure ventilation.
- Free air after laparotomy, abdominal paracentesis and peritoneal dialysis. About 25% of patients still have demonstrable air under the diaphragm after an abdominal operation.
- Sealed subclinical perforation. This may be spontaneous or after gastrointestinal endoscopy including colonoscopy (iatrogenic subclinical perforation).
- Gynaecological causes – tubal insufflation, pelvic examination, douching, etc.
- Pneumatosis cystoides intestinalis.
- Idiopathic – no ascertainable cause.

Pneumoperitoneum can occur after laparoscopic surgery undertaken with CO_2 insufflation but usually clears rapidly within 12 hours owing to the rapid absorption of this gas. Complete aspiration of CO_2 after a laparoscopic operation is important for two reasons: less severe immediate postoperative CO_2 narcosis and reduction of shoulder pain the day after surgery.

In all cases, pneumoperitoneum is of significance only in the presence of symptoms and signs of peritoneal irritation. Otherwise a conservative approach is indicated. An overdistended viscus (e.g. hepatic flexure), adventitial gas shadows, subdiaphragmatic extraperitoneal fat and basal pulmonary collapse may produce radiological appearance simulating free air in the peritoneal cavity – *pseudopneumoperitoneum*.

Haemoperitoneum

The presence of blood in the peritoneal cavity is a major feature of trauma to the intraperitoneal and retroperitoneal organs (Chapter 21). The recent trend in patients with splenic and liver injuries is for conservative treatment unless the patient remains unstable with continued hypovolaemia despite blood transfusion and administration of clotting factors (fresh frozen plasma and platelet transfusion). Crystalloid intravenous therapy is no loner advised in these patients because of the risk of abdominal compartment syndrome.

Other causes of haemoperitoneum include bleeding from the puncture site following percutaneous liver biopsy or other related interventions, and after hepatic resections. In both cases, there is the added problem of abnormal coagulation that often precludes spontaneous arrest. Thus active intervention without any delay is needed in these patients.

The bleeding from puncture site(s) following percutaneous interventions may be substantial if a major intrahepatic vessel is damaged. The initial step during the operation is the application of a vascular clamp across the foramen of Winslow (Pringle's manoeuvre). This is left for 10–20 minutes, after which the clamp is removed and, if the bleeding does not recur, the puncture hole is injected with fibrin glue. Recurrence of the bleeding after release of the vascular clamp indicates damage to a major intrahepatic vessel. There are the two options. The traditional one is to expose the damaged vessel by extending the puncture wound after reapplication of the vascular clamp to the inflow vessels. This is then suture ligated and the clamp released. A simpler technique (if equipment and expertise are available) is localized thermal ablation using a radiofrequency probe. There is a risk in patients with these iatrogenic puncture injuries of the development of an intrahepatic arterioportal fistula, and, for this reason, they should be followed by high-dose CT-based angiography.

Bleeding after hepatic resection necessitates identification of the bleeding points usually in the cut hepatic parenchyma with individual suture ligation. If there are no obvious bleeding points but generalized ooze, argon beam spray coagulation is the best option. If this does not control the oozing, perihepatic gauze packing is indicated. The packs are removed 24–48 hours later.

Haemoperitoneum may arise spontaneously from:

- rupture of hepatic neoplasm
- rupture of splenic and hepatic aneurysms
- severe necrotizing pancreatitis
- peritoneal carcinomatosis.

In clinical practice, haemoperitoneum is seldom a feature of patients with leaking abdominal aneurysm who are admitted alive to hospital as intraperitoneal rupture of the aorta is immediately fatal. The clinical picture of spontaneous haemoperitoneum varies with the underlying cause. When the blood loss is substantive, e.g. rupture of liver tumour or splenic artery aneurysm, the clinical picture is that of severe hypovolaemia and abdominal distension. The hypovolaemia is resistant to volume replacement and the only hope of survival of these patients is immediate surgery with subdiaphragmatic aortic cross-clamping.

In the absence of significant hypovolaemia, the presence of blood in the peritoneal cavity causes some irritation of the peritoneal lining with the development of abdominal signs which, however, vary in intensity and can be marked especially if the haemorrhagic exudate is due to an inflamed organ, e.g. severe pancreatitis, or contains irritant secretions, e.g. bile. Adynamic ileus is invariably present in these cases. The clinical features of bleeding peritoneal carcinomatosis include increase in the ascites and abdominal girth and a hypochromic microcytic anaemia. The diagnosis is confirmed by an abdominal tap. Treatment in these terminally ill patients is entirely supportive.

Tumours of the peritoneum

Secondary peritoneal carcinomatosis

By far the commonest is secondary peritoneal carcinomatosis usually encountered in patients above the age of 40 years. The most common sites of the primary are stomach, breast, pancreas, colon and ovary. As far as tumours of the pancreas, stomach, colon and ovary are concerned, there is now evidence for a stage when viable tumour cells are shed into the peritoneal cavity, before any visible macroscopic appearance of deposits. This stage is identified by lavage cytology with use of specific immunohistochemical stains. In general, exfoliation of tumour cells occurs when the tumour reaches the serosa. In cancer of the colon, 30% of patients have positive lavage cytology at the time of the resection of the primary. In most instances, positive lavage cytology is accompanied by an unfavourable prognosis.

Clinical features

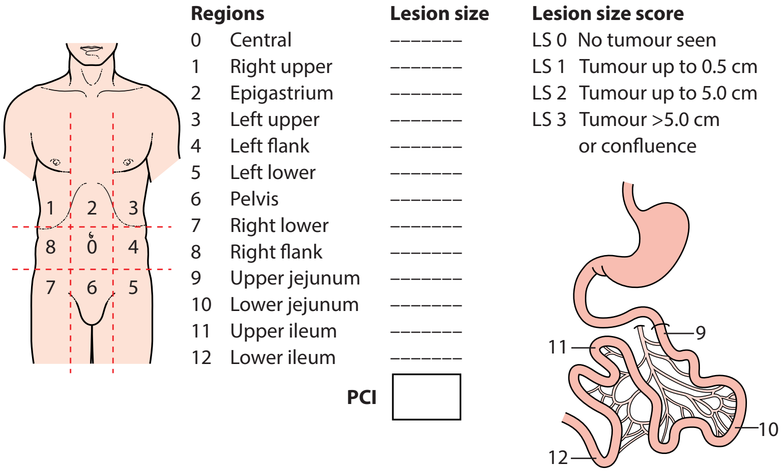
Early peritoneal disease (small seedling deposits) is asymptomatic and can only be detected by investigations. It may or may not be accompanied by deposits in the liver but in most cases the primary tumour has reached the serosa and secondary deposits in the lymph nodes are invariably present. Advanced peritoneal carcinomatosis is in contrast relatively easy to detect on clinical grounds as it is always accompanied by cachexia, anaemia and often malignant ascites which may be blood stained. It signifies advanced terminal disease with median survival of only a few months. Equally it is considered untreatable and the patient is best managed with hospice or supportive care with the necessary palliative therapy to ensure comfort and absence of pain.

Between these two extremes – early peritoneal disease and advanced terminal carcinomatosis – is a spectrum of involvement for which two staging systems are available, and there is one specific Japanese P score for carcinoma of the stomach. The two staging systems commonly used in other cancers/neoplasms are the Gilly score and the peritoneal cancer index (PCI) devised by Sugarbaker *et al.*

The Gilly staging identifies four progressive stages of disease advancement:

- stage 0: no macroscopic peritoneal disease
- stage I: malignant implants <5 mm in diameter localized in one part of the abdomen
- stage II: diffuse to the whole abdomen
- stage III: malignant implants 5 mm to 2 cm
- stage IV: large malignant nodules (>2 cm).

The PCI, described by Sugarbaker *et al.*, is also determined at the time of surgical exploration. However, it provides an estimate of probability of complete cytoreduction and an accurate assessment of survival after cytoreductive surgery and intraperitoneal chemotherapy. The PCI quantitatively combines the distribution of tumour throughout 13 abdominopelvic regions, each with a lesion size score. The abdomen is divided into nine regions, which are numbered in a clockwise direction with 0 at the umbilicus and 1 the space beneath the right hemidiaphragm. Regions 9–12 divide the small bowel into upper



	Regions	**Lesion size**
0	Central	———
1	Right upper	———
2	Epigastrium	———
3	Left upper	———
4	Left flank	———
5	Left lower	———
6	Pelvis	———
7	Right lower	———
8	Right flank	———
9	Upper jejunum	———
10	Lower jejunum	———
11	Upper ileum	———
12	Lower ileum	———
	PCI	☐

Lesion size score	
LS 0	No tumour seen
LS 1	Tumour up to 0.5 cm
LS 2	Tumour up to 5.0 cm
LS 3	Tumour >5.0 cm or confluence

Figure 21.10 Peritoneal cancer index (PCI). Two transverse planes and two sagittal planes are used to divide the abdomen into nine regions, which are numbered in a clockwise direction with 0 being the umbilical and 1 defining the space beneath the right hemidiaphragm. Regions 9–12 divide the small bowel into upper and lower. Lesion size score is determined after complete adhesiolysis enabling inspection of all parietal and visceral peritoneal surfaces. Lesion size score refers to the greatest diameter of tumour implants on the peritoneal surfaces. Disease causing matting of abdominal or pelvic structures together is scored as L-3 even if it is a thin confluence of cancerous implants. [Reproduced from Jacquet, Sugarbaker, in Sugarbaker (ed.) *Peritoneal Carcinomatosis: Principles of Management*. Boston, MA: Kluwer Academic publishers, 1996:359–374.]

and lower jejunum and upper and lower ileum (Figure 21.10). The lesion sizes are then summed for all abdominopelvic regions, giving a score of 0–39.

Diagnosis

The only reliable method for detection of seedling peritoneal tumour deposits is by laparoscopy with laparoscopic contact ultrasonography and this is superior to the standard radiological imaging for establishing the diagnosis. However, the development of special MRI with 3.0T machines and multidetector helical CT scanning has improved the detection of small secondary deposits and some reports seem to suggest that they may replace laparoscopy for stagings The MRI technique involves fat suppression during gadolinium-enhanced MRI after the administration of dilute oral barium solution. This double-contrast MRI enables the detection of carcinomatosis, and tumours <1 cm in diameter in 75–80% of cases.

Treatment

There is no curative treatment for advanced disseminated carcinomatosis; these patients are treated entirely palliatively with hospice care. However, because of the favourable results obtained in patients with pseudomyxoma peritonei (PMP) following cytoreduction surgery and intraperitoneal chemotherapy (see below), this form of therapy has been extended to selected patients with peritoneal metastases secondary to colorectal cancer. The preliminary results have been promising in patients:

- with low-volume, low-grade peritoneal metastases
- with perforated cancers
- in whom definitive surgical cytoreduction is complete.

In these patients hyperthermic intraperitoneal chemotherapy with mitomycin C (35 mg/m^2) results in an actuarial 2 year survival of 59%. Obviously further studies are needed before this treatment is advocated routinely.

Pseudomyxoma peritonei

The term PMP is used to describe rare conditions characterized by the intraperitoneal accumulation of mucoid ascites/mucinous plaques involving the peritoneal surfaces that may or may not contain epithelial and other cells. There has been considerable confusion regarding both the pathology and aetiology of PMP in the past but this is now largely resolved. There are two main clinicopathological types of PMP:

1. *Disseminated peritoneal adenomucinosis* (DPAM): benign peritoneal lesions that consist mainly of extracellular mucin containing few benign mucinous epithelial cells without cytological atypia or mitotic activity. An appendiceal mucinous adenoma (cystadenoma) is present in 50–55%. The condition may also complicate mucocele of the appendix.
2. *Peritoneal mucinous carcinomatosis* (PMCA): malignant peritoneal lesions composed of abundant mucinous epithelium with cytological features of carcinoma.

PMCAs are further classified into:

- PMCA consistent with origin from an appendiceal or intestinal mucinous adenocarcinoma.
- PMCA with features intermediate between DPAM and PMCA or with discordant features despite the presence of at least focal areas of carcinoma in the peritoneal lesions, whether or not the primary site demonstrates an unequivocal carcinoma.

This classification of PMP is important because of its prognostic significance. Thus, with treatment, the age-adjusted 5 year survival rate for patients with DPAM exceeds 80%, as opposed to 38% for patients with PMCA with intermediate or discordant features, and 7% for patients with PMCA.

Pathology

There is some debate on the exact aetiology, as PMP has been reported in association with a variety of neoplasms: appendix, ovary and, rarely, pancreas, colon and urachus.

The immunophenotype (cytokeratin 7 negative; cytokeratin 20 and carcinoembryonic antigen positive) of the lesions of PMP is indicative of a gastrointestinal origin in the vast majority of cases. On very rare occasions, the disease involves the retroperitoneal space either alone or in association with intraperitoneal disease (retroperitoneal PMP). The possibility that some cases of PMP arise from primary transformation of the peritoneal mesothelium into neoplastic mucin-secreting tissue (mucinous metaplasia) has been raised although chromosomal analysis of these tumours does not support this hypothesis. Nonetheless, PMP has to be regarded as a heterogeneous lesion and usually develops from appendiceal (benign or malignant) or ovarian lesions (benign).

PMCA is best considered as a diffuse carcinoma derived from gastrointestinal (usually appendiceal) mucinous cystadenocarcinomas. The cancer cells/mucinous plaques are found at predetermined sites within the abdomen and pelvis and the primary tumour may be small and inconspicuous. The subdiaphragmatic space on both sides, the greater omentum and the pelvis are the sites of maximum disease. Involvement of the intestine is late. Women often have concomitant ovarian mucinous tumours that suggest primary ovarian neoplasia. However, morphological, immunohistochemical and molecular studies in these cases usually indicate that the ovarian tumours are secondary and the PMCA is of intestinal origin. These observations question the existence of a borderline group of mucinous ovarian tumours as a cause for this condition. Thus ovarian tumours exhibiting borderline features should be included in the benign group and designated as atypical proliferative mucinous tumours. Pulmonary parenchymal metastases, although rare, have been reported in patients with PMCA.

In contrast, DPAM represents the benign form of the disease. In these patients the peritoneal implants are usually derived from the extrusion of epithelial cells from an adenoma of the appendix. It is thought that the mucin deposition occurs in accordance with fluid flow and gravitational forces within the peritoneal cavity. The small bowel is not usually involved. Instances of DPAM have been reported in association with mucocele of the appendix (obstructive dilatation of the appendiceal lumen owing to abnormal accumulation of mucus). Mucocele of the appendix has a characteristic CT appearance consisting of spherical or elongated cystic lesions, attached to the wall of the caecum, and may exhibit mural calcification.

Contrary to early reports, cytological examination of the mucin pool/plaques always reveals cells but these vary in count and type (epithelial cells, mesothelial or mesothelial-like cells, histiocytes and fibroblast-like or spindle cells). The epithelial cells may be columnar with mucinous features and may have benign or malignant cytological features. In general, high epithelial cell counts and cytologically malignant cell types are associated with a poor prognosis.

Clinical features

PMP is a rare condition, being reported in approximately two per 10 000 laparotomies. The majority of patients are middle-aged or older patients. Often the primary tumour is slow growing and rarely metastasizes or invades adjacent viscera. In addition to progressive abdominal distension, the condition may present as an acute abdomen with intestinal obstruction. Instances of presentation with spontaneous external fistula in patients with PMP arising from a mucinous carcinoma of the appendix are recorded. PMP has also been reported in association with splenic mucinous epithelial cyst. These may present with splenomegaly before the development of the disease, i.e. presenting feature of PMP, or be detected after the diagnosis or with recurrence of PMP.

When suspected, confirmation of the disease is obtained by fat-suppressed, gadolinium-enhanced, breath-hold MRI after administration of dilute oral barium solution. This carries a higher diagnostic yield than non-enhanced MRI or CT scanning. The characteristic MRI features include thick-walled cysts and septa. Laparoscopy carries the highest diagnostic yield and provides an assessment of the extent of the disease and its pathological type (on biopsy).

Treatment

The treatment is multimodal and consists of:

- aggressive surgical debulking
- intraoperative or postoperative intraperitoneal chemotherapy or systemic combination chemotherapy.

Surgical treatment (usually by laparotomy) entails aggressive surgical evacuation, and resection of the primary and diseased peritoneum. Removal of the appendix is advocated in all cases even if it does not appear to be the site of the primary disease.

Currently most centres favour intraperitoneal chemotherapy with alkylating agents, or 5-fluorouracil (5-FU) and mitomycin C or cisplatin. One report has indicated that hyperthermic (40–42°C) intraperitoneal chemotherapy improves the response rates. Others rely on systemic combination chemotherapy including cisplatin. Radiotherapy is generally regarded as being ineffective in PMP. Long-term survival without recurrence of the disease is well documented in patients with DPAM. By contrast, patients with a high-grade malignant process (PCAM) usually die of the disease within 8 years, but the generally held view is that treatment (if the condition of the patient permits) is the same. In these malignant cases the debulking can be a difficult and time-consuming operation. Others avoid aggressive therapy in view of the high morbidity rates without prospects for significant improvement in survival. This seems to be an unduly pessimistic view. Good results, especially for intermediate PCMAs, can be obtained. They depend on early diagnosis and treatment before large volumes of disease and multiple surgical procedures lead to small bowel entrapment by tumour.

There have been reports of the laparoscopic management of PMP secondary to adenocarcinoma of the appendix. This approach permits thorough exploration of the abdomen, irrigation and aspiration of the thick mucinous material and the instillation of mucolytic agents (5% dextrose solution). Appendicectomy or right hemicolectomy is performed with minimal disturbance of the anterior abdominal wall. The intraperitoneal catheters for chemotherapy are placed through the port sites.

Peritoneal mesothelioma

Mesothelioma is a relatively rare malignant neoplasm arising from the serosal lining of the pleural, peritoneal and pericardial cavities. In many countries there has been a rising incidence of the disease since 1950 owing to exposure to asbestos predominantly in males. Age-specific incidence rates are highest for the older age groups. There is a long latency period between exposure to diagnosis, and this averages 20–30 years. Several studies have shown that these asbestos-related tumours are characterized by balanced chromosomal rearrangements. Asbestos-related mesotheliomas occur most commonly in the pleura, but instances of peritoneal mesotheliomas due to asbestos exposure are well documented and asbestos bodies (mainly amosite) have been documented in the peritoneal cavity, the most common sites being the mesentery and omentum, usually in patients with heavy-fibre burdens in lung tissue.

Peritoneal mesotheliomas can develop, albeit rarely, in the absence of exposure to asbestos. Perhaps the most important is malignant mesothelioma after radiation therapy for Hodgkin disease. In these patients, the mesothelioma arises in the field of prior radiotherapy. There is a rare primary variant that occurs in young females, and instances of mesotheliomas have been reported in patients with familial Mediterranean peritonitis.

Pathology

Malignant mesotheliomas are pleural, pericardial or peritoneal neoplasms usually associated with asbestos exposure. Since mesothelial cells are biphasic, they may give rise to epithelial and sarcomatous malignant mesotheliomas. In addition, benign 'atypical proliferations of mesothelial cells' may occur. The separation of mesothelial hyperplasia from early malignant mesothelioma can be very difficult on histological examination. These lesions are termed 'atypical mesothelial hyperplasia'. Immunostaining for epithelial membrane antigen and the quantification of silver-stained nucleolar organizer regions can be useful in differentiating benign from malignant histological sections of pleural and peritoneal biopsies. There have been recent reports of simian virus 40 DNA large T antigen sequences in both pleural and peritoneal epithelial malignant mesotheliomas but not in sarcomatous lesions. Malignant mesothelioma tends to be a diffuse lesion and, when associated with asbestos exposure, is classified as an industrial disease. Pathological confirmation of asbestos aetiology is necessary for these patients to obtain compensation.

Benign mesotheliomas comprise two macroscopic types: fibrous and multicystic. The fibrous type forms well-encapsulated solid tumours composed of spindle-shaped cells. The multicystic variety has a marked tendency for local recurrence after resection. Both types are associated with long survival after surgical resection.

In females, the histological distinction between epithelial peritoneal mesothelioma and ovarian papillary serous carcinoma diffusely involving the peritoneum may be difficult even with immunohistochemistry and usually requires examination of the ultrastructural features by electron microscopy. To compound a difficult problem further, rare instances of primary mesothelioma in young women have been documented and some of these are associated with highly elevated serum levels of cancer antigen 125. This makes the differentiation even more difficult since this is the typical marker for epithelial serous tumours from the ovary.

Clinical features

Currently only 2.2 cases of malignant mesotheliomas per million population per year are diagnosed. The incidence of the disease continues to increase because of the long latency period, despite the fact that exposure to asbestos in industry has virtually ceased since 1970. Malignant mesotheliomas are encountered mainly in men. The mean age at first diagnosis of malignant mesotheliomas is approximately 59 years; women on average are 4 years younger at presentation. The most frequent initial symptoms of malignant mesotheliomas include fatigue, abdominal pain, anorexia, marked weight loss and abdominal distension caused by intractable ascites. Clinical presentation as fever of unknown origin is exceptional, but, when it occurs, it signifies a very aggressive tumour with a bad prognosis. Both benign and malignant disease may present with an abdominal mass or intestinal obstruction or gastric outlet obstruction. More uncommon presentations include dysphagia secondary to pseudoachalasia, chronic pancreatitis and regional lymphadenopathy. Overall, the mean survival time ranges from about 3 months to 8 years. Prolonged survival is only encountered with benign mesotheliomas.

The current standard non-invasive imaging modality for staging of malignant mesothelioma is CT. However, CT does not determine resectability. Laparoscopy using a multiport technique with diaphragmatic, peritoneal and abdominal wall biopsies provides more accurate staging. The ascitic fluid is examined for malignant cells.

Treatment

This entails surgical exploration with excision (benign lesions) or tumour debulking followed by intraperitoneal chemotherapy. Details of the intraperitoneal chemotherapy vary. Good results have been reported with continuous hyperthermic peritoneal perfusion with cisplatin-based regimens. In phase II studies, this combination results in a median progression-free survival of 26 months, and an overall 2 year survival of 80% (for all mesotheliomas including the multicystic variants). The morbidity is acceptable (24%). Retreatment after initial response can result in a second long-term response.

Some advocate two-stage peritoneal chemotherapy for patients with primary peritoneal mesothelioma. In stage I, patients undergo cytoreductive surgery and placement of an intraperitoneal infusion catheter, through which intraperitoneal chemotherapy is administered for 4 months. In stage II, a second laparotomy with debulking of residual disease and placement of perfusion intraperitoneal catheters is carried out. High-dose intraperitoneal hyperthermic (40–42°C) chemotherapy using a disposable perfusion circuit is then administered. The reported experience with this two-stage treatment is limited and currently there is no evidence that it gives superior results.

Malignant mesotheliomas are generally regarded as being unresponsive to systemic chemotherapy. However, dramatic regression of the disease has been reported following systemic chemotherapy with gemcitabine. The activity of gemcitabine in malignant mesothelioma has been confirmed by phase II studies. There is also evidence suggesting better response rates when gemcitabine is combined with cisplatin. Other regimens that have shown activity are based on doxorubicin and cisplatin.

Special forms of intestinal obstruction

Postoperative intestinal obstruction

It is often incorrectly assumed that a persistent intestinal obstruction (beyond 3–4 days) after abdominal surgery is due to a protracted adynamic ileus. While there are some specific operations in which this is the case, e.g. intestinal intubation with a Baker's tube or complex remedial reconstructive surgery on the gastrointestinal tract, in the majority of patients in whom the obstruction is delayed a mechanical cause is likely to be present. Early obstruction (during the first 5 days) is usually due to non-strangulating causes, e.g. anastomotic oedema, adhesive fibrinous matting with distension and kinking of the intestinal loops. As the obstruction is often incomplete, active surgical intervention is rarely necessary and the majority settle with conservative management. By contrast, later postoperative obstructions (occurring at or persisting beyond 7 days of surgery) are usually caused by organized bands or abscesses and may be strangulating in nature. Laparotomy is therefore necessary for these late or prolonged postoperative intestinal obstructions.

Bolus obstruction

Intraluminal bolus obstruction, usually of the small intestine, may be caused by tricho- and phytobezoars, gallstones or a mass of worms (*Ascaris lumbricoides*). The last is usually encountered in children and is often precipitated by antihelminthic therapy.

Phytobezoars

These are firm masses of undigested fruit or vegetable fibres that can cause gastric or small bowel obstruction. The predisposing factors include:

- ingestion of large amounts of high-fibre foods
- inadequate mastication
- previous gastric surgery producing hypo- or anacidity and loss of the antral pump mechanism.

Often the phytobezoars that form after gastric surgery consist of orange pith.

Patients with gastric bezoars present with epigastric pain, loss of appetite and weight, and episodes of distension and vomiting. The condition is usually diagnosed at endoscopy. Intestinal bezoars present with mechanical small bowel obstruction. Gastric bezoars are multiple in 17% of cases and intestinal bezoars in 4%.

Gastric bezoars may be treated conservatively by cellulase enzymatic digestion (300 mL of 0.5% cellulase solution instilled 4 hourly via a nasogastric tube for 2 days). Alternatively piecemeal removal may be carried out endoscopically but this may be difficult. Laparotomy is needed for patients presenting with small bowel obstruction. Sometimes it is possible to knead the bolus into the caecum. When this is not successful, a gastrotomy is performed.

Intestinal pseudo-obstruction

The term 'pseudo-obstruction' is used to describe obstruction of the small or large intestine in the absence of a mechanical cause or acute intra-abdominal disease. The term covers a variety of syndromes which result from damage to the myenteric plexus (neuropathy) or smooth muscle abnormality (myopathy), or both. Small intestinal and colonic pseudo-obstruction are best discussed separately.

Small intestinal pseudo-obstruction

This condition may be primary (idiopathic) or secondary. Familial *hollow visceral myopathy,* which is included in the primary category, is a particularly severe disorder which involves the smooth musculature of the oesophagus, entire gastrointestinal tract including the colon and often the urinary bladder. The secondary variety results from a neuropathy/myopathy induced by certain systemic disorders or drug misuse (excess phenothiazine administration, laxative abuse). The disorders most commonly associated with the development of secondary small intestinal pseudo-obstruction are:

- diabetes mellitus
- scleroderma
- progressive systemic sclerosis
- acute intermittent porphyria
- hypothyroidism
- Chagas disease.

It has also been reported as a complication of sclerotherapy for oesophageal varices. When the underlying abnormality is a neuropathy (e.g. diabetes mellitus), the pattern of intestinal motor activity is abnormal with derangements of the myoelectrical migratory complexes, absence of any normal activity or disorganized non-propulsive hypermotility. By contrast, in myopathic conditions (e.g. hypothyroidism), the pattern of motor activity is normal but the intensity of the contractile activity is reduced.

The clinical picture is that of recurrent episodes of subacute intestinal obstruction with colicky abdominal pain, vomiting and distension. The treatment entails correction of any underlying disorder whenever this is possible (e.g. hypothyroidism). Intestinal prokinetics, e.g. metoclopramide and domperidone or cisapride, are sometimes beneficial, especially the last. Cisapride acts by increasing the local concentration of acetylcholine in the intestinal smooth musculature but this drug has been withdrawn from clinical practice because of severe side effects. The synthetic peptide ceruletide, which has to be administered intravenously or intramuscularly, is also beneficial, particularly during acute episodes. Intravenous erythromycin may be effective in some patients. Replacement therapy is necessary in patients with hypothyroidism.

Colonic pseudo-obstruction

Two types are recognized: acute and chronic. The acute condition was first described by Ogilvie in 1948 and is sometimes referred to as *Ogilvie syndrome*. It consists of a selective dilatation of the caecum and proximal colon with a sharp cut-off usually at the splenic flexure, and less frequently the hepatic flexure and sigmoid, suggestive of mechanical obstruction. It is best considered as a localized form of adynamic ileus, which usually develops in patients with major pre-existing, non-intestinal conditions requiring hospitalization, e.g. major surgery, severe trauma, sepsis, myocardial infarction, severe renal and respiratory disease. The aetiology remains unknown although administration of drugs that impair colonic motility and air swallowing are thought to be contributory factors. The defect appears to be in the smooth muscle or in one of the intestinal control mechanisms, and slow wave activity (electrical control activity) has been reported to be absent in these patients.

The clinical picture is dominated by abdominal distension often without vomiting. The patient complains of increasing discomfort and may develop cramp-like abdominal pain. Radiography shows features of colonic obstruction with caecal distension (7.5–22.0 cm). Caecal perforation is a well-recognized complication and is likely to occur if the radiological size of the caecum exceeds 12 cm.

The standard treatment is by colonoscopic decompression, which is successful in the majority of patients. Recurrence of the condition is encountered in 20% of patients. These are treated by further colonoscopy when a Baker's intestinal tube is inserted transanally into the caecum, or the colonoscope left *in situ* for 2 hours, thereby maintaining decompression. Surgical intervention with tube caecostomy is undertaken when colonoscopic treatment fails.

Chronic colonic pseudo-obstruction may be primary or secondary. They latter may result from a motility disorder, diabetes mellitus, hypothyroidism, malignancy, psychosis or drug and laxative abuse. This is covered elsewhere (Chapter 30).

Intussusception

Intussusception is a telescoping of a segment of intestine into an adjacent one (Figure 21.11). The condition is encountered most commonly in childhood with a peak incidence at 4 months. Intussusception may, however, be encountered in the adult, in which case a precipitating lesion that initiates the intussusception (the lead point) is usually present, e.g. intestinal polyp or submucous lipoma. By contrast, in infants and children some 70–95% are classed as idiopathic (no lead point) and an associated illness, such as urinary tract infection or gastroenteritis, is encountered in 30%. It is often assumed that hyperplasia of the lymphoid patches in the terminal ileum secondary to common disease in infancy may be involved in the initiation of idiopathic intussusception in this age group. A definite seasonal incidence is observed in infants and children with clear peaks in the spring and summer. This is suggestive of viral infections being the cause of the intestinal lymphoid hyperplasia.

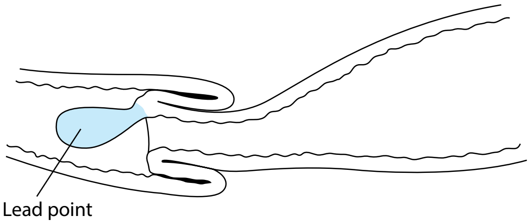


Figure 21.11 Components of an intussusception.

Intussusception is anatomically defined as ileoileal, ileocaecal and ileocolic depending on the site and extent of the telescoping observed by the time of diagnosis. The condition is a strangulating type of intestinal obstruction and, if treatment is delayed, ischaemic necrosis of the involved bowel segments and peritonitis are inevitable.

The clinical features include the sudden onset of vomiting, cramp-like abdominal pain and rectal bleeding. An abdominal mass is palpable in 55–60% of cases. Children with intussusception associated with a lead point (Meckel's diverticulum, polyp, duplication, Henoch–Schönlein purpura, suture line, appendix, tumour) are usually older than the idiopathic cases.

In the absence of peritonitis and intestinal obstruction, the initial treatment is by hydrostatic barium enema, which is successful in 50% and is followed by a recurrence rate of 6%. Intussusception caused by a lead point is not likely to reduce with a hydrostatic barium enema and usually requires surgical intervention. Operative treatment is necessary when the hydrostatic reduction is incomplete, uncertain or contraindicated (peritonitis, intestinal obstruction). Operative reduction is usually possible, although the viability of the bowel may be compromised after reduction and resection is therefore necessary. In adults, it has to be assumed that a lead point is present and its nature determined by subsequent investigations after recovery, unless of course the lesion is palpable through the bowel wall after reduction.

Volvulus

Volvulus is a twist or rotation of a loop of intestine about its mesenteric attachments. It is therefore a sudden obstruction of the closed loop variety if the rotation is complete (360°) and ischaemia or total vascular occlusion may be present by the time of diagnosis.

The condition may involve the stomach (Chapter 23), small and large intestine. Two intestinal varieties are described: primary and secondary. Primary volvulus results from malrotation of the gut or congenital excessive mobility from loose fixation or long mesenteric attachments (volvulus of the mid-gut in the neonate). The more common secondary volvulus is due to rotation of a loop of small intestine around an adhesions/band, an ileostomy or colostomy. Caecal and sigmoid volvulus are considered elsewhere (Chapter 30).

Intraperitoneal adhesions

Adhesions following abdominal and pelvic surgery are important in view of their morbidity and frequent hospital readmissions. They occur after elective surgery and following peritonitis from any cause. The morbidity spectrum of intraperitoneal adhesions includes:

- chronic pain
- acute and subacute intestinal obstruction

- secondary female infertility
- increased operating time for subsequent operations
- risk (20%) of iatrogenic bowel injuries during subsequent operations.

Pathology of adhesions

The causes of intraperitoneal adhesions are:

- open abdominal or pelvic surgery
- individual susceptibility
- ischaemic areas: anastomotic sites, reperitonealization of raw areas
- foreign bodies: talc, starch granules, gauze lint, cellulose, etc.
- peritonitis from any cause
- inflammatory bowel disease: Crohn' disease
- radiation enteritis
- sclerosing peritonitis: usually drug induced, e.g. certain beta-blockers.

The most common category is postoperative adhesions. There is undoubtedly an individual susceptibility as some patients are prone to adhesions and others not, even after major operations, but the genetics and molecular biological basis for this predisposition to adhesion formation are not known. Talc is no longer a problem. Surgical gloves either incorporate a hydrogel polymer or are powdered with epichlorohydrinated cornstarch. Gauze lint is still important and there is evidence that it is responsible for 25% of cases of intraperitoneal granuloma formation caused by implanted foreign bodies. There is some evidence that some postoperative adhesions develop on a background of ischaemia in the region of surgically constructed anastomoses and following attempts at reperitonealization of raw areas. These should be left unsutured, when they are rapidly filled with an inflammatory exudate and thereafter covered by a new serosa derived from free-floating peritoneal macrophages.

The exact pathogenesis of intra-abdominal adhesions is not fully understood but trauma and exposure of the peritoneum and the subsequent biochemical and cellular response to repair this injury are involved. Fibrin deposition is the initiating step as this is essential for mesothelial repair. Normally there is a fine balance between resurfacing of denuded areas by mesothelial cells and fibrinolysis that removes the initial fibrin structural framework. Disturbance of this fine balance between mesothelial repair and fibrinolysis by ischaemia and other factors is thought to be responsible for adhesion formation after abdominal surgery. As the fibrinolytic activity is compromised, the fibrin matrix persists, becomes invaded by fibroblasts and gradually matures into fibrous bands, usually within 5–7 days. Adhesions may be parietal (between viscera and abdominal walls and diaphragm) or visceral (between bowel loops and solid organs). Usually both are present in the individual case.

Another factor that is thought to be important in the development of adhesions is disruption of the naturally occurring surface active phospholipid (SAPL) barrier, which normally keeps the peritoneal membrane lubricated and separates the visceral from the parietal peritoneum. The component lipids of this SAPL barrier, which protects other surfaces such as the alveolar membrane, are dipalmitoylphosphatidylcholine (DPPC) and unsaturated phosphatidylglycerol (PG). Three-dimensional studies of the molecular configuration have shown that the SAPL barrier is composed of multiple layers (lamellar bodies). The efficacy of SAPL lamellar bodies is due to their ability to bind to the mesothelial epithelium within the microvilli to form molecular structures that keep the surface well lubricated and dewetted. The protective and lubricating properties of the SAPL barrier are compromised during surgery and may account for the development of adhesions.

There is no established classification of adhesions. They can be filmy, dense or string-like, and of variable spread within the peritoneal cavity, but are usually centred on the operative site and the parietal access wound. From the symptomatic and pathological viewpoint, parietal adhesions (those binding the intra-abdominal contents to the parieties) are more important than interloop/visceral adhesions.

Adhesions after elective surgery

Although estimates of the incidence of adhesions vary (40–97%), all the reports indicate that the problem is substantial and results in both a significant morbidity and major added healthcare costs. In the USA over 400 000 operations were performed for lysis of adhesions in 1993 and the total costs for treating patients with complications and symptoms caused by adhesions has been estimated at $1.2 billion annually. In one retrospective cohort study based on validated data of patients undergoing open abdominal or pelvic surgery (29 760) from the Scottish National Health Service database in 1986, followed up for 10 years, 34.6% were readmitted a mean of 2.1 times during the study period for a disorder directly or possibly related to adhesions, or for abdominal or pelvic surgery that could be potentially complicated by adhesions. Six per cent of all readmission during the study period were directly due to adhesions and the majority of these (97%) required an operation during the readmission. Most of the adhesion-related events are encountered during the first year after surgery (22% of all operative adhesiolysis performed in the first 12 months) but the incidence of adhesion-related morbidity and readmissions did not decline during the 10 year period (Figure 21.12).

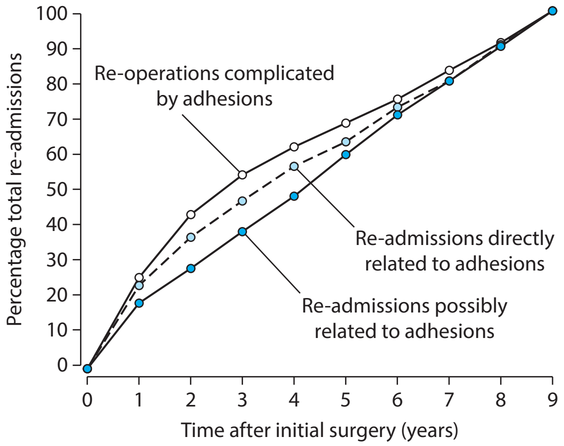


Figure 21.12 Adhesion-related hospital readmissions after abdominal and pelvic surgery: a retrospective cohort study. (Reproduced by kind permission from Ellis, *Lancet* 1999;**353**:1476–1480.)

Complications of adhesions requiring readmission to hospital are much more frequent after surgery on the small bowel and colon than after pelvic gynaecological surgery involving the female reproductive organs. Small bowel obstruction developed in 25% of patients after total colectomy with ileoanal pouch reconstruction in one large reported series of 1005 patients. Approximately 70% of all hospital admissions for small bowel obstruction are due to adhesions.

Clinical features

Patients with intraperitoneal adhesions may develop chronic symptoms or present acutely with intestinal obstruction.

Chronic adhesion syndromes

The most common is abdominal pain with or without nausea in the absence of abdominal signs, except perhaps minimal tenderness. The pain distribution is variable and is not necessarily beneath the abdominal wall scar. These are the most difficult patients to evaluate and they usually undergo a series of investigations, often repeated before coming to either laparotomy or laparoscopy. The outcome in terms of pain relief after adhesiolysis is uncertain; some patients obtain pain relief (possibly a placebo effect) but usually the pain recurs. In other patients the pain is episodic and colicky in nature, accompanied by nausea, audible excessive bowel sounds and abdominal distension lasting for variable periods. These attacks terminate in diarrhoea, which relieves both the pain and the abdominal distension. In these patients the symptoms are due to recurrent attacks of subacute small bowel obstruction and, indeed, these patients may develop acute obstruction at any time. These patients often lose weight from malabsorption due to bacterial overgrowth. This group undoubtedly benefits from surgery (see below).

Acute adhesive intestinal obstruction

This is the most common reason for emergency admission/readmission to hospital. The obstruction is almost always in the small bowel although the level of obstruction varies. Colonic obstruction due to adhesions is rare and other causes should be considered. The severity of vomiting and the extent of abdominal distension vary with the level of the obstruction. The condition is readily diagnosed by the clinical symptoms and signs and abdominal plain films (erect and supine). The possibility of strangulation should be suspected if the pain is constant, there is obvious tenderness on palpation and a marked leucocytosis. Early recourse to surgical intervention may prevent infarction of the intestine.

Treatment of adhesion-related complications

Prophylactic

In the first instance good surgical technique is important. In this context, the specific points include:

- delicate handling of tissues and organs
- avoidance of spillage of visceral contents during surgery
- minimizing operative blood loss
- protection of exposed viscera from drying
- avoidance of closure of parietal peritoneal defects – this induces ischaemia at the suture line
- washing the operative region and peritoneal cavity with isotonic saline at the end of the operation.

There is good evidence that laparoscopic surgery is followed by a significantly reduced incidence of postoperative adhesions and this effect is probably related to maintenance of the 'milieu intérieur' of the coelomic cavity during the operation, although other factors may be involved.

Even with the best operative techniques, elective surgery on the small intestine and colon and emergency surgery for acute peritonitis is a significant risk factor and there is an increasing argument for specific prophylaxis after these operations. A variety of drugs [anticoagulants, dextrans of various molecular sizes, antihistamines, non-steroidal anti-inflammatory drugs (NSAIDs), povidone, streptokinase etc.] have been investigated as antiadhesion agents but none have been shown to be effective as chemoprophylactic agents against adhesion formation.

There is to date only one proven prophylactic treatment that should be considered in these patients. This consists of a bioresorbable membrane based on sodium hyaluronate (Seprafilm) which is placed on the front of the intestinal loops/greater omentum separating them from the parietal wound. The use of this membrane has been shown to be entirely safe (no detectable adverse effects) and in a prospective randomized trial of patients undergoing colectomy and ileal pouch anal anastomosis significantly reduced the incidence of adhesions (15% versus 58%).

Based on the role of the SAPL barrier in peritoneal lubrication and its disruption by open surgery, a synthetic mixture of surface-active phospholipids (DPPC and PG – proprietary name ALEC) has been used in experimental studies and shown to reduce adhesions by 70%. The product is available in powder form (identical to surfactant used in neonatal respiratory problems) and is jet sprinkled into the peritoneal cavity at the end of the procedure. Clinical studies are ongoing in patients undergoing emergency colonic operations with ALEC but to date there are no published results.

Another antiadhesion agent of interest is a 4% icodextrin solution (Adept) used extensively in peritoneal dialysis. As these patients do not form peritoneal adhesions, icodextrin has been studied experimentally and in one clinical trial against Ringer's solution in gynaecological laparoscopic surgery. In both instances it reduced the incidence, severity and extent of adhesions. However, its efficacy in general surgical practice remains unproven. It has several advantages. These include proven non-toxicity, low cost and ease of administration, i.e. 1 litre of the solution is instilled in the peritoneal cavity at the end of the operation. Icodextrin solution is contraindicated in patients with known allergy to starch-based polymers and in patients with maltose or isomaltose intolerance.

Treatment of patients with symptomatic/complicated adhesions

Patients with chronic symptoms

The patients with a clinical picture of subacute small bowel obstruction should undergo surgical adhesiolysis. This may be undertaken by the open approach or laparoscopically depending on local expertise. Following adhesiolysis, insertion of Seprafilm between the contents and the parieties is advocated. The insertion of Seprafilm may prove difficult if the operation is conducted by the laparoscopic approach. The decision for adhesiolysis is difficult in patients with chronic symptoms in the absence of definitive abdominal signs. If the surgeon decides to intervene, the laparoscopic approach is preferable to the open approach in these patients, as the likelihood is that the symptoms will recur after a variable period and the patient will then be insistent on further surgery. The author had a patient referred for remedial surgery to reconstruct the abdominal wall who, by the age of 49 years, had had 38 operations. The alternative to Seprafilm is Adept solution, particularly if the operation is conducted laparoscopically.

Patient with acute small bowel obstruction

In the absence of signs and symptoms of strangulation, these patients are managed conservatively with nasogastric decompression and fluid and electrolyte replacement in the first instance. This conservative management is usually successful but if there is no improvement within 12–24 hours of admission, emergency laparotomy and adhesiolysis is undertaken. These operations should be performed under prophylactic antibiotic cover, and, if for any reason the bowel is injured during the operation or resection of an intestinal segment is performed, a full 5 day antibiotic course to cover Gram-negative organisms (anaerobes and aerobes) is necessary. The laparoscopic approach is used by some surgeons in these patients but is technically difficult in view of the distended oedematous intestinal loops and may carry a higher risk of iatrogenic bowel injury than the open conventional approach. In some instances, because of dense adhesions, the bowel may be injured during the mobilization. Inadvertent enterotomy during adhesiolysis occurs in 15–20% and increases the postoperative morbidity and urgent relaparotomy rate. These patients are also more likely to require intensive care support and parenteral nutrition and have significantly longer postoperative stay in hospital. The risk factors for inadvertent enterotomy during adhesiolysis are:

- obesity
- three or more previous laparotomies
- adhesiolysis in lower abdomen and pelvis
- old age.

If the inadvertent enterotomy is ragged and extensive, it is better to resect this injured segment; otherwise, it is oversewn. The insertion of Seprafilm between the intestinal loops/omentum and the parietal wound is advisable in all patients undergoing adhesiolysis for intestinal obstruction (irrespective of iatrogenic bowel injury). Patients with features of strangulating obstruction at or soon after admission (especially those with constant pain and marked leucocytosis) require immediate surgery following resuscitation.

Patients with recurrent episodes of small bowel obstruction

These patients pose real problems in surgical management, as the relative efficacy of the various surgical operations used for this condition has never been adequately assessed by prospective randomized studies. The operations available include:

- adhesiolysis alone
- adhesiolysis with Seprafilm
- Noble intestinal plication
- Child–Phillips placation
- Baker intestinal intubation.

To date there is no evidence of any clear superiority of one procedure and there are no data on the efficacy of Seprafilm in reducing the obstructive episodes after adhesiolysis, but this seems a sensible addition to adhesiolysis especially as its use does not incur any adverse effects.

The Noble intestinal plication (Figure 21.13) involves suturing the small intestinal loops with serosal sutures so that the small bowel coils become fixed in gentle curves when adhesions reform. The procedure is rarely performed nowadays because it is time-consuming and carries an appreciable morbidity (perforation, fistula, peritonitis). Furthermore, patients complain of chronic abdominal pain and remain subject to recurrent attacks of intestinal obstruction.

In the Child–Phillips operation of transmesenteric plication (Figure 21.14) after adhesiolysis, the small intestine and mesentery are placed in an orderly fashion. Thereafter, the intestinal

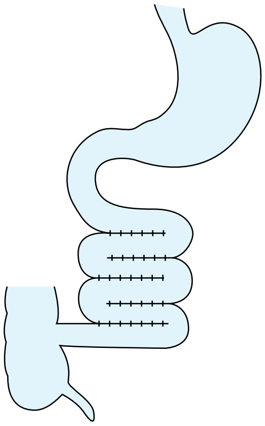


Figure 21.13 Noble's plication operation. The small intestinal loops are sutured together in smooth curves by means of seromuscular sutures.

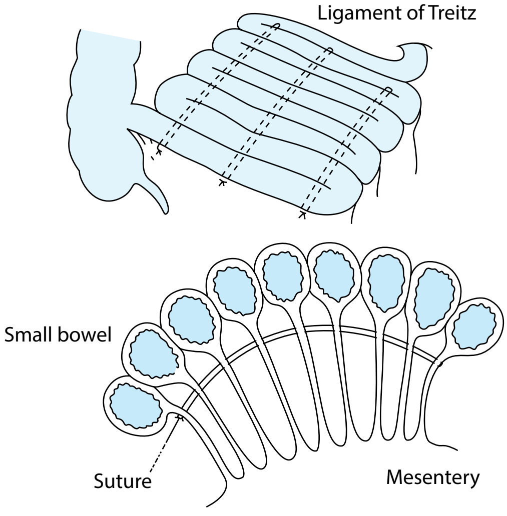


Figure 21.14 The Child–Phillips operation of transmesenteric plication for recurrent small intestinal obstruction due to adhesions.

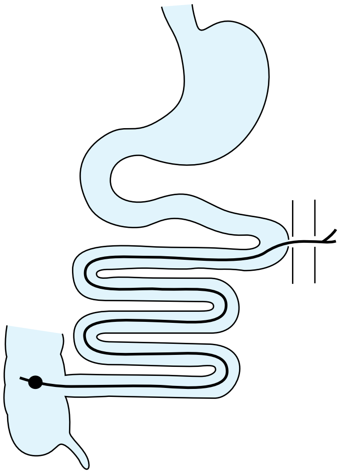


Figure 21.15 Intestinal intubation for recurrent small bowel obstruction using a Baker's tube inserted via a Witzel jejunostomy.

coils are fixed in position by means of transmesenteric sutures or some 3–5 mm away from the mesenteric border of the intestine, avoiding the terminal branches of the mesenteric arcade. The results of a number of retrospective reports with this procedure for recurrent small bowel obstruction due to adhesions have been good in terms of relative freedom from further episodes of intestinal obstruction.

The intraluminal tube or long intestinal 'stent' tube was first proposed by White as a simple method of achieving gentle curves rather than allowing the bowel to develop acute angulations when adhesions reform. Baker introduced the jejunostomy tube and advocated its insertion through either a purse-string suture in the upper jejunum or a Witzel jejunostomy (Figure 21.15). The tube has an inflatable balloon near its tip, which facilitates passage down the small intestine into the caecum.

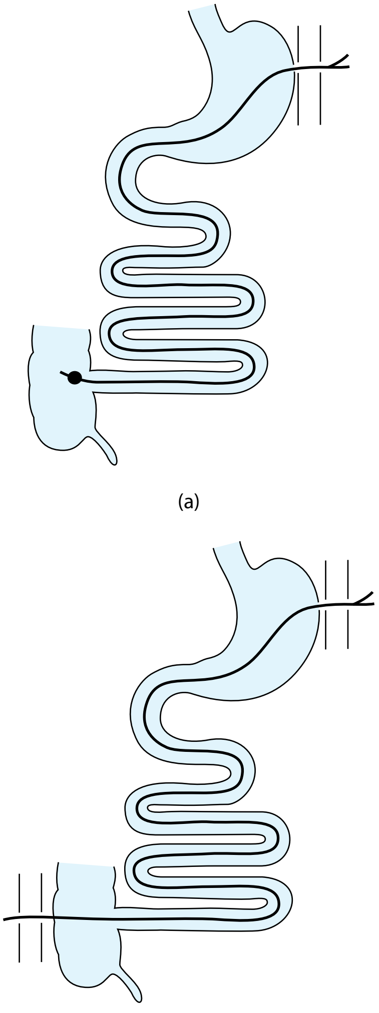
Modifications of the technique include gastrostomy stent plication with and without tube exit-caecostomy (Figure 21.16). The latter is unnecessary and increases the risk of infection. One disadvantage of the intestinal intubation method is the long period of ileus after the operation. Reports on the efficacy of intestinal intubation in preventing recurrent intestinal obstruction have been conflicting. Some have found the procedure beneficial and have extended its use as a prophylactic measure in the treatment of patients with generalized peritonitis and following major abdominal surgical procedures. Other reports indicate that intestinal intubation is inferior to adhesiolysis alone.

Desmoid tumours (aggressive fibromatosis)

The pathological disorder underlying the development of these lesions is *aggressive fibromatosis*. Desmoid tumours are considered as locally aggressive but non-metastasizing tumours. This is not strictly true as instances of peritoneal dissemination are documented but are uncommon. These tumours exhibit a marked tendency to recur after surgical excision, lead to considerable morbidity and not uncommonly contribute to death of the patients. Aggressive neurofibromatosis may arise sporadically or in association with familial adenomatous polyposis (FAP) and Gardner syndrome.

Pathology

Aggressive fibromatosis (myofibromatosis in children) is a fibroproliferative disorder of fascial and musculoaponeurotic structures that most commonly affects the trunk and extremities but variants can affect the intra-abdominal mesenteries and the parietal pleura. Instances of desmoids developing in the breasts are well documented but rare. The pathological process consists of a monoclonal proliferation of spindle (fibrocyte like) cells that are locally invasive with a tendency to recurrence but very rarely metastasize, although instances of peritoneal dissemination have been reported in patients with mesenteric desmoids. Some tumours arise sporadically but a subgroup is associated with FAP. -Catenin mutations and -catenin dysregulation have been documented in both sporadic aggressive fibromatosis and that occurring in association



(a)

(b)

Figure 21.16 (a) Gastrostomy with long intestinal tube plication for recurrent small bowel obstruction. (b) Modification of the technique with gastrotomy and tube exit-caecostomy. This technique is not generally recommended as the exit-caecostomy is considered unnecessary and increases the risk of infection.

with FAP, itself caused by germ-line mutations in the adenomatous polyposis coli (*APC*) gene. In desmoid tumours, one of the two mutations usually occurs distal to the second β-catenin binding/degradation repeat of the gene (3' to codon 1399). Catenin and catenin-binding genes are known to be associated with neoplastic processes. Independent predictors of increased risk of developing desmoid are (1) germ-line mutation distal to codon 1399, (2) family history of gastrointestinal disease and (3) a strong family history of desmoid tumours.

In genetically normal individuals (sporadic cases), at least two somatic mutations must occur in both alleles of the *APC* gene (which normally acts as a tumour suppressor gene). In FAP syndrome patients, one *APC* germ-line gene is already mutated and, therefore, only one new somatic mutation is required in the opposite *APC* gene for a desmoid to develop. It is known that the *APC* gene is involved in the regulation of the cellular level of β-catenin, which has a cell membrane function (member of the adherens junctions) and binds transcription factors in the nucleus, thereby activating transcription (mediator in Wingless signalling). Cellular β-catenin protein level is elevated in all desmoid tumours (sporadic and those associated with FAP). The demonstration of mutations in two mediators in the Wnt-APC-β-catenin pathway implicates β-catenin stabilization as the key factor in the pathogenesis of aggressive fibromatosis. It appears that truncating *APC* gene mutations (337 bp insertion of an *AluI* sequence into codon 1526 of the *APC* gene) give aggressive fibromatosis cells a proliferative advantage through elevated cellular levels of β-catenin protein. What is interesting is the observation that the higher level of β-catenin protein compared with surrounding normal tissues is associated with normal levels of β-catenin mRNA. This suggests that the elevated β-catenin is due to degradation at a lower rate than normal tissues.

Desmoid tumours are oestrogen receptor and progesterone receptor negative. Despite this, antioestrogens (tamoxifen and toremifene) and NSAIDs (sulindac and diclofenac) have been shown to induce objective response rates in both adult patients and children with unresectable or recurrent desmoid tumours and the effect of the combination is synergistic, i.e. increased efficacy compared with either agent alone. Most desmoid tumours express the proliferating cell nuclear antigen, although this has not been shown to influence outcome.

Clinical features

Desmoid tumours are rare with a reported incidence of two to four per million population. They can occur at any age from childhood to old age but there is an established female preponderance and the majority of sporadic cases occur in women of childbearing age, in whom they usually appear during or after pregnancy. Infantile/juvenile tumours tend to be very aggressive and involve proliferation of fibrous and muscle layers (myofibromatosis). The mesenteric variant constitutes about 10% of all desmoid tumours and accounts for 50–70% of lesions developing in patients with FAP. Although the trunk and extremities are the most common sites of external desmoids, lesions can occur anywhere – neck, chest, hands, etc.

Table 21.1 Staging system for intra-abdominal desmoid tumours in familial adenomatous polyposis

Stage	Description
I	Asymptomatic, <10 cm maximum diameter, and not growing
II	Mildly symptomatic, <10 cm maximum diameter, and not growing
III	Moderately symptomatic or bowel/ureteric obstruction, or 10–20 cm, or slowly growing
IV	Severely symptomatic, or >20 cm, or rapidly growing

Mildly symptomatic, sensation of mass, pain, but no restrictions.
Moderately symptomatic, sensation of mass, pain; restrictive but not hospitalized.
Severely symptomatic, sensation of mass, pain; restrictive and hospitalized.

The problem with clinical management of desmoid lesions relates to their rarity and the many different therapies available. Standardization is needed so that prospective studies can be planned. In 2005, Church *et al.* proposed a staging system for intra-abdominal desmoid tumours in FAP, based on tumour size and behaviour. This has since been used in sporadic cases as well. The system proposed is simple but useful (Table 21.1).

The majority of tumours developing in patients with FAP do so after surgery and 50–70% are intra-abdominal (mesenteric variant) and 30–50% external in the abdominal wall. Mesenteric desmoids have a significantly worse prognosis than abdominal wall tumours and cause recurrent intestinal or ureteric obstruction, which often contributes to the death of the patient. By contrast abdominal wall desmoids do not cause death or significant morbidity, although recurrence is common after excision.

CT or MRI are used for diagnosis and to determine the extent of the tumour and its relationship/involvement of adjacent structures, to detect recurrence after surgery and to assess response to non-surgical treatment, especially radiotherapy.

Treatment

In view of the rarity of the condition the treatment is not standardized. There treatment modalities are:

- surgical excision
- radiotherapy
- combined antioestrogen and anti-inflammatory drug therapy
- chemotherapy
- newer modalities: isolated limb perfusion and immunotherapy with interferon.

Surgical excision

This remains the most effective, although recurrence rates are high (average 30% at 5 years), recurrence-free survival is good (80–75%) at 5 years for primary extremity and trunk desmoid tumours. Local recurrence after surgical excision is not related to any specific prognostic factors (age, sex, depth and size of tumour, and positive resection margins). Thus attempts at wide excision resulting in unnecessary morbidity may not prevent local recurrence. Surgical excision of these lesions should remove macroscopic disease with preservation of function and structure when possible, as residual microscopic disease does not significantly compromise the 5 year disease-free or overall survival. Conservative surgery is especially indicated in children, in whom it should be accompanied by adjuvant therapy (see below). In limb lesions, amputation should be reserved only for patients in whom the disease or repeated resections have resulted in a non-functional or chronically painful extremity.

Radiotherapy

Radiation is an effective modality for desmoid tumours, either alone or as an adjuvant to resection. The main indication for radiotherapy is unresectable disease. Postoperative radiation is not recommended for patients with negative resection margins. Patients with positive margins should receive 50 Gy of postoperative radiation. Unresectable tumours are irradiated with a dose of approximately 56 Gy with a 75% expectation of local control. The radiation dose correlates with the incidence of complications. Thus doses of 56 Gy or less produce a 5% complication rate at 15 years compared with 30% with higher doses.

In 30% of juvenile patients with aggressive fibromatosis primary complete resection is not feasible. In this age group, radiotherapy as primary treatment is indicated only if complete tumour resection is not feasible without mutilation. In general, radiotherapy should be used as a last resort in children with skeletal immaturity because of the risk of growth disturbance, contracture and secondary malignancy.

Medical therapy

The most common medical treatment used is combined antioestrogen (usually tamoxifen) and NSAID (sulindac) therapy, which can be used at any age but is indicated in patients with unresectable disease/recurrence after radiotherapy, especially children, in whom it is recommended prior to resort to chemotherapy. Reasonable response rates (up to 30%) and disease control are well documented. More recently, the tyrosine kinase inhibitor imatinib has been used with promising results in some patients whose tumours express one of imatinib's molecular targets, platelet-derived growth factor receptor. Early results with antifibrotic agents (pirfenidone) appear promising but the reported experience is very limited.

Chemotherapy

Chemotherapy is used in patients with unresectable tumours, and in those with tumours that do not respond to tamoxifen and/or sulindac. Low-dose chemotherapy with methotrexate and vinblastine produce high response rates, particularly in children. The combination of methotrexate and vinorelbine produces similar response rates but with less nerve damage. Liposomal doxorubicin appears to be more effective (because of increased tumour uptake) than doxorubicin. Low-dose chemotherapy is often given for a year, then stopped for assessment of response by CT or MRI; it requires several months for detection of improvement.

High-dose chemotherapy treatment is more active but is accompanied by increased toxicity; for this reason, it is administered for a shorter duration, usually 6 months in patients unresponsive to other therapies or with very rapidly growing symptomatic tumours. The most commonly used regimen is doxorubicin combined with dacarbazine.

Newer modalities

Isolated limb perfusion with tumour necrosis factor and melphalan has been used with some success as a limb preservation treatment in patients with recurrent desmoids and significant symptoms who would otherwise require mutilating surgery for control of the disease. However, the reported experience, though promising (40% regression rate), is limited. Also interferon-induced remission in aggressive fibromatosis of the lower extremity has been reported. Other new therapies include intralesional sclerosant therapy with acetic acid or absolute alcohol and radiofrequency thermal ablation, but the long-term benefit of both is not known.

Disorders of the retroperitoneum

Retroperitoneal fibrotic disorders

There is a group of fibrotic disorders of the pneumoperitoneum that are closely related and considered to be distinct from aggressive fibromatosis (desmoids). They include:

- secondary retroperitoneal fibrosis: drug-induced, extravasation of urine, desmoplastic response to a variety of tumours
- idiopathic retroperitoneal fibrosis (Ormond disease)
- inflammatory aortic aneurysm
- sclerosing mesenteritis
- multifocal retroperitoneal fibrosclerosis.

All are rare. The secondary type accounts for about 30% of cases. The drug-induced type is associated with the administration of methysergide for migraine but other drugs (methyldopa, phenacetin, etc.) have been implicated.

Idiopathic retroperitoneal fibrosis (Ormond disease)

The condition accounts for 70% of cases and results in the development of a flat, grey–white plaque of tissue that is found in the lower lumbar region and then spreads laterally and upwards to encase the common iliac vessels, vena cava and the aorta. Rarely, it may extend upwards above the renal arteries and become contiguous with a similar process in the mediastinum, or forwards into the small bowel mesentery or mesocolon. It may very rarely also involve the bile duct, duodenum and pancreas. The histological picture varies from an active one which shows inflammatory cells and small blood vessels surrounding bundles of collagen fibres, to a more mature process that is relatively avascular and acellular with patches of calcification.

The aetiology of idiopathic retroperitoneal fibrosis is unknown. It has been attributed in the past to lymphatic and venous obstruction and lymphangitis. Leakage of blood or urine from trauma into the retroperitoneal space causing fibrosis is classed as secondary retroperitoneal fibrosis. The majority of idiopathic cases are regarded as part of an obscure collagen disorder. A possible genetic predisposition for idiopathic retroperitoneal fibrosis is suggested by the presence of the human leucocyte antigen B27 immunophenotype in 44% of cases. The association in some patients with inflammatory aortic aneurysm is now established (see below). In these patients it appears that the fibrosis is caused by a chronic inflammatory or autoimmune response to antigens leaking into the retroperitoneum from atheromatous plaques in the aorta or common iliac arteries.

The average age at presentation is 50–55 years but the disease is documented in young individuals and elderly patients (aged 44–71 years). There is a male predominance in reported cases. Clinical presentation may be insidious with a variable picture but the most common presentation is with hypertension, obstructive uropathy (single or bilateral) and renal failure. A nagging, ill-defined, low back pain is common and there may be evidence of venous obstruction as shown by swelling of the scrotum and legs. Uncommonly, the patient may present with claudication. Examination reveals anaemia, leucocytosis and an increased sedimentation rate. Biochemical investigation may show a degree of renal failure. The intravenous urogram reveals a characteristic picture consisting of hydronephrosis and medial displacement of the ureters with gross irregularity.

Patients in renal failure will require cystoscopy and ureteric stenting to relieve the obstructive uropathy and permit recovery of the renal function. If stenting is unsuccessful, percutaneous nephrostomy is necessary to improve renal function before definitive treatment. This consists of surgical ureterolysis. The operation is carried out by a transabdominal approach, and the freed ureter(s) are separated from the fibrous tissue by vascularized omental interposition and placed intra- or extraperitoneally. This operation gives the best results (even without subsequent corticosteroid therapy) with complete resolution of the symptoms and long-term successful alleviation of ureteric obstruction in 100% of patients.

Many drugs have been shown to effective in idiopathic retroperitoneal fibrosis. These include corticosteroid, azathioprine, cyclophosphamide, methotrexate, tamoxifen and mycophenolate mofetil. Tamoxifen and corticosteroids seem to be the most effective combination and are indicated in patients with continuous activity of the disease and to prevent recurrence after surgery.

Inflammatory aneurysm

Inflammatory aneurysm (aortitis) accounts for 4% of abdominal aneurysms and has an established association with retroperitoneal fibrosis. Following surgical treatment of the aneurysm, the inflammatory fibrosis resolves completely in only 23%, improves in 35% but remains static in the remainder and can then lead to involvement of the ureter(s) or intestine. The characteristic perivascular distribution of idiopathic retroperitoneal fibrosis is in support of the hypothesis that the disease is an immune-mediated response to leaking antigens from severe atherosclerosis. It is recommended that patients with inflammatory aneurysms should be followed up and receive medical therapy with corticosteroids and tamoxifen if the fibrosis does not resolve, or progresses.

Sclerosing mesenteritis

This is a very rare idiopathic benign mesenteric lesion and only a few cases have been reported. It is characterized by fat necrosis, fibrosis and chronic inflammation and has been recently reported in HIV-positive patients. It presents with recurrent abdominal pain and weight loss and tends to form masses that involve the pancreas and the small bowel mesentery. The condition responds to tamoxifen and corticosteroids.

Multifocal idiopathic fibrosclerosis

Multifocal idiopathic fibrosclerosis is a rare syndrome characterized by exuberant fibrosis involving many organs/systems – mediastinal fibrosis, retroperitoneal fibrosis, orbital pseudotumour, Riedel's thyroiditis and sclerosing cholangitis. Patients may also have Dupuytren's contractures, lymphoid hyperplasia, Peyronie disease, vasculitis, testicular fibrosis and pachymeningitis.

Retroperitoneal swellings

The term retroperitoneal tumour is usually confined to lesions arising from tissues (muscles, fat, fibrous tissue, lymph nodes, nerves and developmental remnants) of this compartment but excluding origin from the retroperitoneal organs (pancreas, kidneys and ureters and adrenal glands). Retroperitoneal swellings may be cystic or solid, benign or malignant.

Benign cystic lesions

The cystic swellings are usually benign and the majority are discovered accidentally. They include:

- cystic lesions arising from developmental (Wolffian) remnants of the urogenital tract; these are situated near one or other kidney
- mesenteric cysts (intra- and retroperitoneal)
- teratomatous and dermoid cysts
- abdominal cystic lymphangiomas (lymphogenous cysts)
- parasitic cysts.

Abdominal lymphangiomas

Lymphangiomas are benign lesions of vascular origin that show lymphatic differentiation. The vast majority (95%) occur in the neck and axillary regions with the remainder being located in the abdomen and retroperitoneum, lung and mediastinum. On ultrasound scanning, lymphangiomas appear as multilocular cystic masses that are anechoic but may contain echogenic debris. Intravenous contrast-enhanced CT may show enhancement of the cyst wall and any septa. On MRI, the signal pattern resembles that of fluid: low signal intensity on T_1 weighted imaging and high signal intensity on T_2 weighted images, although haemorrhage or infection in the cystic lesions may alter both the CT attenuation and MRI signal pattern giving a more solid appearance.

In one of the largest series of 107 abdominal lymphomas Levy *et al.* reported the following distribution of abdominal cystic lymphangioma:

- mesenteric, 54%
- retroperitoneal, 10%
- splenic, 8%
- colonic, 7%
- small bowel, 7%

- pancreatic, 5%
- rare site: hepatic, biliary, renal, adrenal, bladder, ovary.

Retroperitoneal lymphangiomas tend to form large elongated lesions. Large mesenteric lymphangiomas may simulate ascites but can be differentiated from it by the presence of septa, compression on adjacent small bowel loops and absence of free fluid in the paracolic gutters and subhepatic spaces.

Lymphangiomas may also involve the gastrointestinal tract: oesophagus, stomach and especially the small intestine and colorectum. Most intestinal lymphangiomas form mural masses usually found incidentally at endoscopy or on radiological studies. The aetiology of intestinal lymphangiomas is variously attributed to intramural lymphatic obstruction, disturbed endothelial permeability, inflammation, congenital absence of lymphatics and ageing of the bowel wall. Barium contrast studies show smoothly marginated mural masses that deform on external abdominal compression. Endoscopic sonography and CT show evidence of cystic mass in the intestinal wall.

Clinical features

Abdominal cystic lymphangiomas are rare and usually present in infants and children, mean age 5 years (range 3 months–8 years) and indeed some are diagnosed on the prenatal ultrasound examination. They may, however, present in adult patients. The majority (90%) of childhood lesions are intraperitoneal (mesentery and gastrointestinal tract) but some (10%) are located in the retroperitoneal space. The symptomatology varies but the most common presentation is with abdominal pain. A palpable tumour is found in 30% of patients. In children, abdominal lymphangiomas may present with complications such as intestinal obstruction (including volvulus), infection and sudden increased pain and swelling caused by intracystic haemorrhage. The diagnosis is established by abdominal ultrasonography, CT and MRI.

In adults, symptoms of abdominal cystic lymphangiomas have usually been present for several months to years, and the documented sites include the pancreas, spleen and retroperitoneum. Thus, compared with children, a significantly higher proportion involve retroperitoneal structures and acute presentation is rare.

Treatment

Treatment is with surgical excision whenever possible and is indicated only for symptomatic or complicated disease. Resection has to be complete to prevent recurrence.

The treatment of colonic symptomatic intestinal lymphangiomas is by colonoscopic endoscopic deroofing of the cysts or polypectomy if pedunculated and small to intermediate in size, but large lesions usually require colonic resection.

Generalized lymphangiomatosis

The diffuse condition of generalized lymphangiomatosis is a rare congenital disorder which consists of multifocal lymphagiomatous lesions. It usually presents in childhood and involves multiple organs including the lung, liver, spleen, bone and skin. Phangiomatosis (generalized lymphangiomatosis) is caused by maldevelopment of the lymphatic systems during the intrauterine period. The condition is most commonly diagnosed in children (65%) and has an equal sex incidence. Cross-sectional imaging with CT and MRI is important for the diagnosis because of the widespread nature of the condition. There are some isolated reports of an association with other mesenchymal dysplasias, such as *Maffucci syndrome* and diaphyseal aclasis. Generalized lymphangiomatosis covers a large spectrum of abnormalities and affects multiple organs. Cases with localized lesions carry a better prognosis but the prognosis of generalized lymphangiomatosis also depends on the extent of the disease. Skeletal involvement is often present when the bones most frequently involved are the tibia, humerus, ilium, skull, mandible, vertebrae and small bones of the hands. The bony lesions are osteolytic with multiple septae and these lesions can cause complications: pathological fracture, joint deformity and severe pain. Some patients have extensive mediastinal, pleural and peritoneal involvement with cystic fluid collections. The prognosis is bad in patients with combined mediastinal and bone lesions. Splenic lymphangiomatosis shows characteristic findings on CT imaging: enlarged spleen containing multiple well-circumscribed cystic lesions that usually show no contrast enhancement. Although rare, subcutaneous nodules, are well documented in the literature.

Treatment

Treatment of generalized lymphangiomatosis is usually conservative. Surgical intervention is indicated only for complications such as pathological fractures and symptomatic/complicated disease. Bisphosphonate medication is used to prevent bone loss in skeletal lymphangiomatosis. Abdominal complications are treated by laparotomy with resection of the diseased organ involved in the complication.

Solid tumours

The majority of solid retroperitoneal tumours are malignant. The benign tumours, which constitute 20%, include:

- lipomas
- neurofibromas, neurilemmomas
- leiomyomas
- extra-adrenal chromaffinomas (phaeochromocytomas) – may be malignant
- retroperitoneal mucinous cystadenomas
- haemangiopericytoma.

The *Costello syndrome* is characterized by mental retardation and the occurrence of benign tumours (usually papillomata) but instances of children with the Costello syndrome who develop retroperitoneal embryonal rhabdomyosarcoma are documented. Adult retroperitoneal tumours of nervous origin account for 20% of the retroperitoneal tumours in this age group. The majority are males (sex ratio 10:3). The main symptom is the abdominal pain (84.6% of all cases) and all the tumours usually reach a large size by the time of presentation. Benign solid lesions such as retroperitoneal neurilemmomas show up as well-demarcated round or oval masses on CT. They may exhibit heterogeneous contrast enhancement, no enhancement indicative of cyst formation and tumour calcification.

Paragangliomas are tumours of embryological origin from the neural crest and can be found in any location along the aorta and/or in association to the sympathetic chain. These tumours can be non-functioning or secreting [catecholamines – usually norepinephrine (noradrenaline)] and the secretory state influences the clinical presentation. These secreting neuroendocrine tumours are responsible for 0.1–0.5% cases of hypertension. Only 20% of paragangliomas are catecholamine secreting, and cause a syndrome similar to that of phaeochromocytoma. Paraganglionomas may be multiple and malignant. The malignant types recur locally after excision and spread by the bloodstream predominantly to bones and lungs. Immunohistochemical staining of paraganglionomas may exhibit type I cells (chromogranin A and neurone-specific enolase), which have no prognostic significance, or type II cells (S100 protein), the presence of which is associated with a good prognosis. Some paraganglionomas are familial (hereditary paraganglioma) and are associated with germ-line mutations in the von Hippel–Lindau disease tumour suppressor gene and the *ret* proto-oncogene.

Both benign (mucinous cystadenomas) and malignant (mucinous cystadenocarcinomas) can occur as primary tumours of the retroperitoneum. The immunological staining characteristics of malignant lesions indicate that these tumours have patterns similar to ovarian mucinous tumours. This genotypic similarity with ovarian mucinous tumours may indicate similar mechanisms in their histogenesis.

Haemangiopericytomas are rare vascular benign tumours derived from pericytes and can develop in the retroperitoneal space where they usually present as very bulky but otherwise clinically silent tumours. Tumours as large as 30 cm have been documented. Treatment is by surgical excision.

Although not strictly a primary retroperitoneal tumour as angiomyolipoma (a mesenchymal tumour) usually originates from the kidney, it is considered here for a number of reasons. In the first instance, its pleomorphic appearance and involvement of regional lymph nodes may simulate malignancy, although the lesion is benign. Second, angiomyolipomas are very common (approximately 40%) in patients with tuberous sclerosis and are then usually small, bilateral and asymptomatic.

The important clinical feature of angiomyolipomas is spontaneous rupture and retroperitoneal haemorrhage and this usually occurs with larger tumours. Although the tumour has specific ultrasonographic and CT appearances, histological examination is necessary for establishing the diagnosis. The management varies with size. Thus small asymptomatic lesions should be followed up with sequential CT scans and removed by enucleation or partial nephrectomy only if they reach a size >4.0 cm. Emergency surgery is needed for patients with massive bleeding, when it is usually very difficult to preserve the kidney.

Overtly malignant retroperitoneal tumours are of diverse origin:

- lymphomas
- congenital neuroblastoma
- soft-tissue sarcomas: fibrosarcoma, liposarcoma, rhabdomyosarcoma, neurofibrosarcoma and other malignant nerve cell tumours
- neoplasms arising from the urogenital ridge.

Congenital neuroblastoma

Congenital neuroblastoma is an embryonal malignancy of the sympathetic nervous system arising from neuroblasts. By definition, *congenital neuroblastoma* is defined as a neuroblastoma diagnosed within a month of birth, and is of types: (1) *fetal* and (2) *neonatal* neuroblastoma. The vast majority of fetal neuroblastomas (90%) originate in the adrenal glands. They are identified by fetal ultrasound examination usually in the third trimester (36 weeks' gestation). In most cases the tumour undergoes maturation and regression, and prognosis for the fetal type is excellent with conservative management. The prenatal detection necessitates delivery in a perinatal centre so that neonatal surgery, if needed, can be performed.

Neonatal neuroblastomas have a much better prognosis than those presenting in older children with cure rates of 90%. The adrenal glands are the site of origin in only 45% of neonatal neuroblastomas and in 60% of infants metastases, most frequently to the liver, are present at the time of diagnosis and cause massive hepatomegaly with respiratory impairment. In contrast 70–80% of children older than 18 months present with widespread metastatic disease, usually to the lymph nodes, liver, bone and bone marrow and <50% can be cured, even with aggressive treatment.

The cause of neuroblastoma is unknown. The vast majority of tumours arise sporadically without any family history. A small minority (1%) have a family history of neuroblastoma and these patients usually present at an earlier age or with multiple primary tumours. Neuroblastoma can occur in association with other disorders linked to abnormal development of neural crest tissues, e.g. Hirschsprung disease and central congenital hypoventilation syndrome. These cases have a germ-line mutation in *PHOX2B*, a homeobox gene that acts as a regulator for the development of the autonomic nervous system.

Pathology

The tumour is composed of undifferentiated neuroblasts, which histologically appear as small, round, blue cell tumours. A peculiar arrangement of dense nests of cells in a fibrovascular matrix –'pseudorosettes' – is seen in 15–50% of tumours – these are neuroblasts surrounding eosinophilic neuritic processes. These neuroblast tumour cells are small and uniform cells with scanty cytoplasm and hyperchromatic nuclei. Formation of neuritic processes is a pathognomonic feature of neuroblastoma cells. Immunohistochemical staining with neuron-specific enolase, chromogranin, synaptophysin and S-100 stains are usually positive. Electron microscopy is used to detect ultrastructural features of the tumour: neurofilaments, neurotubules, synaptic vessels and dense core granules.

Chromosomal and molecular markers

The most important biological marker is *MYCN,* an oncogene that is overexpressed in approximately 25% of neuroblastomas through the amplification of the distal arm of chromosome 2. Patients whose tumours exhibit *MYCN* amplification have aggressive rapidly growing disease and a poor prognosis. In contrast, expression of the *H-ras* oncogene correlates with a better prognosis and lower stage of the disease at presentation. Deletion of the short arm of chromosome 1 is the most common chromosomal abnormality present in neuroblastoma and, when present, confers a poor prognosis. DNA index is another useful parameter that correlates with therapeutic response in infants with neuroblastoma: tumours which exhibit hyperdiploidy (DNA index >1) have a good therapeutic response to chemotherapy with cyclophosphamide and doxorubicin, whereas infants with tumours exhibiting a DNA index of 1 are less responsive, requiring more aggressive combination chemotherapy.

The neurotrophin receptor gene products *TrkA*, *TrkB* and *TrkC*, are tyrosine kinases that code for receptors of members of the nerve growth factor family. In infants <12 months, a high expression of *TrkA* correlates with a good prognosis. In contrast, *TrkB*, which is more commonly expressed in tumours with *MYCN* amplification, is associated with poor prognosis. Other biological markers associated with poor prognosis include increased levels of telomerase RNA and lack of expression of glycoprotein CD44 on the tumour cell surface. P-glycoprotein and multidrug resistance protein are two proteins expressed in neuroblastoma that confer multidrug resistance in some cancers.

Histopathological classification (Shimada)

This is based on (1) the degree of neuroblast differentiation, (2) extent of Schwannian stroma (stroma rich or stroma poor), (3) cellular proliferation (mitosis–karyorrhexis index; MKI), (4) nodular pattern and (5) age. It has been confirmed to be very useful in differentiating patients into those with favourable histology indicative of good prognosis and those with unfavourable histology indicative of aggressive disease.

Favourable groups include:

- patients of any age with stroma-rich tumours without a nodular pattern
- patients <18 months with stroma-poor tumours, MKI <200/5000 (200 karyorrhectic cells per 5000 cells scanned) and differentiated or undifferentiated neuroblasts
- patients <60 months with stroma-poor tumours, MKI <100/5000 and well-differentiated tumour cells.

Unfavourable groups include:

- patients of any age with stroma-rich tumours and a nodular pattern
- patients of any age with stroma-poor tumours, undifferentiated/differentiated neuroblasts, MKI >200/5000
- patients >18 months with stroma-poor tumours, undifferentiated neuroblasts, MKI >100/5000
- patients >18 months with stroma-poor tumours, differentiated neuroblasts, MKI of 100–200/5000
- patients >60 months with stroma-poor tumours, differentiated neuroblasts, MKI <100.

Clinical features

Neuroblastoma constitutes 6–10% of all childhood cancers, and 15% of cancer deaths in children. The highest incidence is in the first year of life, and some cases are congenital. Although the age range is broad, including older children and even adults, only 10% of neuroblastomas occur in individuals >5 years old. A large European study reported that <2% of 4000 neuroblastoma patients were over 18 years old. International reports indicate high incidence rates in the rich Western countries (Europe and North America), and lower incidence in low-income/developing countries (Africa, Asia and Latin America). Males have a slightly higher incidence of neuroblastoma than females (male to female ratio 1.2:1). The age distribution at the time of diagnosis is: 40% of patients are <1 year, 35% are aged between 1 and 2 years and 25% are >2 years when diagnosed. According to SEER, incidence decreases every consecutive year up to age 10 years, after which the disease is rare.

The majority (90%) of neuroblastomas are secretory with elevated levels of catecholamines or their metabolites (dopamine, homovanillic acid or vanillylmandelic acid) being present in the blood or urine or blood. The common symptoms include abdominal pain, vomiting, weight loss, anorexia, fatigue and bone pain. Chronic diarrhoea is a rare presenting symptom secondary to autonomous secretion of vasoactive intestinal peptide by the tumour. Overall, 50% of patients present with advanced stage disease with osseous and bone marrow deposits. Thus, the most common presentation is with bone pain and a limp. Some patients may develop unexplained fever, irritability and periorbital ecchymosis secondary to metastatic disease to the orbits. Pathological fractures are commonplace in patients with bony metastases.

In 70% of patients, the primary tumour is intra-abdominal and hence an asymptomatic abdominal mass, usually discovered by the parents, is a common early manifestation. In time the mass becomes symptomatic, with the symptoms depending on the proximity of the tumour to vital structures. Tumours arising from the paraspinal sympathetic ganglia can grow through the spinal foramina into the spinal canal and impinge on the spinal cord producing a spectrum of neurological manifestations ranging from weakness and limping to paralysis and even bladder/bowel dysfunction. In a small proportion of infants <6 months, neuroblastoma presents with a small primary tumour and metastatic disease to the liver, skin and bone marrow (stage 4S). In neonates, the skin lesions may be confused with congenital rubella. When the skin involvement is extensive, the term ‘blueberry muffin baby’ is used to describe the striking appearance. Rarely (2%) patients present with opsoclonus and myoclonus – a paraneoplastic syndrome characterized by the presence of myoclonic jerking and random eye movements. These patients often have localized disease and a good long-term prognosis. The neurological abnormalities may, however, persist.

Clinical staging (International Neuroblastoma Staging System)

This is as follows:

- Stage 1: localized tumour with complete gross excision, microscopic residual disease, or both. Histologically negative ipsilateral lymph nodes (nodes attached to the primary may be positive for tumour).

- Stage 2A: localized tumour with incomplete gross resection. Representative ipsilateral non-adherent lymph nodes microscopically negative for tumour.
- Stage 2B: localized tumour, complete gross excision, or both with positive ipsilateral non-adherent lymph nodes. Enlarged contralateral lymph nodes, which are histologically negative.
- Stage 3: unresectable unilateral tumour infiltrating across the midline, regional lymph node involvement, or both. Alternatively, localized unilateral tumour with contralateral regional lymph node involvement.
- Stage 4: any primary tumour with dissemination to distant lymph nodes, bone, bone marrow, liver, skin and/or other organs (except as defined for stage 4S).
- Stage 4S: localized primary tumour (as for stages 1, 2A or 2B) with dissemination limited to the skin, liver and/or bone marrow (<10% involvement). Limited to infants.

More recently the International Neuroblastoma Task Force has proposed the International Neuroblastoma Risk Group (INRG) classification system as retrospective studies have revealed high survival rates of children in the 12–18 month old age group previously categorized as high risk. In this group of children the tumour does not have amplification of the *myc* gene.

The new INRG risk assignment will classify neuroblastoma at diagnosis based on a new International Neuroblastoma Risk Group Staging System (INRGSS) is as follows:

- Stage L1: localized disease without image-defined risk factors.
- Stage L2: localized disease with image-defined risk factors.
- Stage M: metastatic disease.
- Stage MS: metastatic disease 'special', in which MS is equivalent to stage 4S.

The risk stratification is based on the new INRGSS staging system: age (dichotomized at 18 months), tumour grade, N-*myc* amplification, unbalanced 11q aberration and ploidy into four pretreatment risk groups: very low, low, intermediate and high risk.

Treatment

Most centres now follow the INRG treatment strategies with different treatment regimens for (1) low-risk, (2) intermediate-risk and (3) high-risk groups.

Treatment strategy for low-risk patients

Patients with localized resectable tumours (stage 1) have excellent event-free survival (EFS) rates with surgical excision of tumour only. Adjuvant chemotherapy is not indicated in these patients but is used if patients develop recurrent disease. The overall survival rate is high (>95%). Similar management is indicated in patients with stage 2A/2B disease if they fall in a low-risk category and have non-amplified *myc* tumours, regardless of age, histology or ploidy. However, patients with stage 2A/2B disease with *myc*-amplified tumours are considered high risk regardless of age and histology. Most patients with 4S disease (non-*myc*-amplified tumours, favourable histology, hyperdiploid tumours in infants <12 months) are also considered a low-risk group as most experience spontaneous regression. Thus, management consists of observation or surgery alone. Chemotherapy is used to control life-threatening situations such as respiratory distress from a large liver.

Treatment strategy for intermediate-risk group

Surgery and multiagent chemotherapy are used for intermediate-risk patients. Intermediate-risk patients include (1) children <18 months with stage 3 or 4 disease and favourable biology (non-*myc*-amplified tumours, regardless of histology and DNA index). These patients are treated with the four cytotoxic agents active against neuroblastoma (cyclophosphamide, doxorubicin, carboplatin, etoposide) using four cycles, six cycles or eight cycles, depending on histology and DNA index and response to treatment. In these patients, surgery can be performed either at the time of diagnosis or after combination chemotherapy. Radiotherapy is considered for residual disease after chemotherapy and surgery. However, this remains controversial.

Treatment strategy for high-risk group

These patients seem to require aggressive treatment with multiagent chemotherapy, surgery and radiotherapy, followed by consolidation with high-dose chemotherapy and peripheral blood stem cell rescue. Current protocols involve four phases of therapy: *induction, local control, myeloablative consolidation* and *treatment of minimal residual disease*. *Induction therapy* consists of multiagent chemotherapy with alkylating agents, platinum, and anthracyclines and topoisomerase II inhibitors. *Local control* is by surgical resection of primary tumour and radiotherapy to the primary tumour site. Primary tumours often become resectable after induction chemotherapy. *Myeloablative consolidation therapy* has been shown to improve EFS in high-risk patients and is administered using etoposide, carboplatin and melphalan. Most centres now recommend the use of peripheral blood stem cell support over bone marrow during myeloablative consolidation therapy. *Treatment of minimal residual disease* with biological agents improves survival. To date, most experience has been with 13-*cis*-retinoic acid. But, increasingly, other biological agents are being added to 13-*cis*-retinoic acid, e.g. CCG-389, interleukin 2, granulocyte macrophage colony-stimulating factor and the chimeric anti-GD2 (gangliosidase) antibody.

Surgical treatment

Surgery has an important role but forms part of multidisciplinary treatment. Surgical resection is curative in patients with localized disease. For patients with regional or metastatic disease it is often used for second-look surgery after chemotherapy, when residual disease is resected or debulked as completely as possible without sacrificing major organ function.

Prognosis

Between 20% and 30% of high-risk cases do not respond adequately to induction high-dose chemotherapy. Most long-term survivors are patients with low- or intermediate-risk disease. Even so many survivors have long-term adverse effects from the treatment. Thus, survivors of intermediate- and high-risk disease often have hearing loss, growth reduction, thyroid function disorders and learning difficulties in addition to a greater risk of secondary cancers in high-risk disease patients.

Retroperitoneal sarcomas

Retroperitoneal sarcomas constitute only 10–20% of all sarcomas and have an overall incidence of 0.3–0.4% per 100 000 of the population with a peak incidence in the fifth decade of life. The most frequent cell types are liposarcoma, leiomyosarcoma and malignant fibrous histiocytoma (MFH). Recently, however, retroperitoneal MFH has been disputed as a separate entity by immunohistochemistry studies, which have indicated that most retroperitoneal MFHs are dedifferentiated liposarcomas. Despite advances in molecular oncology, the aetiology in most soft-tissue sarcomas remains elusive. At the time of presentation, these tumours are usually large owing to their slow, relatively asymptomatic, growth over a period of many months (Figure 21.17). Local recurrence is very common (33–86%) after surgical resection and local failure is usually the cause of death as distant metastases are rare (up to 30%) in reported series.

Pathology

Retroperitoneal liposarcomas constitute the commonest type of retroperitoneal sarcomas and carry a worse prognosis than liposarcomas of the trunk and extremities largely because of the late and advanced stage at presentation. Liposarcomas are mesenchymal neoplasms containing both atypical adipocytes and lipoblasts in an adipose tissue stroma. They exhibit variable biological aggression in terms of local invasion and metastatic risk depending on the degree of differentiation. Irrespective of site of origin, liposarcomas are classified into four main histological subtypes: (1) well differentiated, (2) myxoid, (3) round cell (low risk) and (4) dedifferentiated (high risk/grade). The dedifferentiated type carries a bad prognosis in terms of recurrence and survival. The mitotic index of retroperitoneal sarcomas reflects the grade as follows:

- sarcomas with no or one mitosis per 10 high-power fields (HPFs) (low risk)
- sarcomas with one to nine mitoses per 10 HPFs (intermediate risk/grade)
- sarcomas with more than nine mitoses per 10 HPFs (high risk/grade).

Some liposarcomas exhibit on histology peculiar meningothelial-like whorls and metaplastic bone in the whorls or immediate vicinity The meningothelial-like whorls represent a mesenchymal proliferation that may undergo myofibroblastic or osteoblastic differentiation in liposarcoma. The meningothelial whorls are thought to represent an early sign of dedifferentiation of liposarcomas. CT has been shown to reliably distinguish between well-differentiated and dedifferentiated retroperitoneal liposarcomas based on percentage fat content, focal nodular/water density, ground glass opacities and hypervascularity. Dedifferentiated tumours are more likely to be infiltrative and hypervascular with areas of necrosis. The CT-based radiological diagnosis of a well-differentiated lesion is 100% accurate, and avoids the need for preoperative biopsy in these patients.

Leiomyosarcomas are also more aggressive when found in the retroperitoneum. On histology, these tumours consists of cells with cigar-shaped nuclei, and stain positive for actin on immunohistochemistry. Leiomyosarcomas arise either from large vascular structures or from non-vascular tissues, with patients with primary vascular lesions exhibiting a worse prognosis in terms of overall median and metastasis-free survival (2.1 years and 0.25 years, respectively) than those with leiomyosarcomas of non-vascular origin (7 years and 9.6 years, respectively).

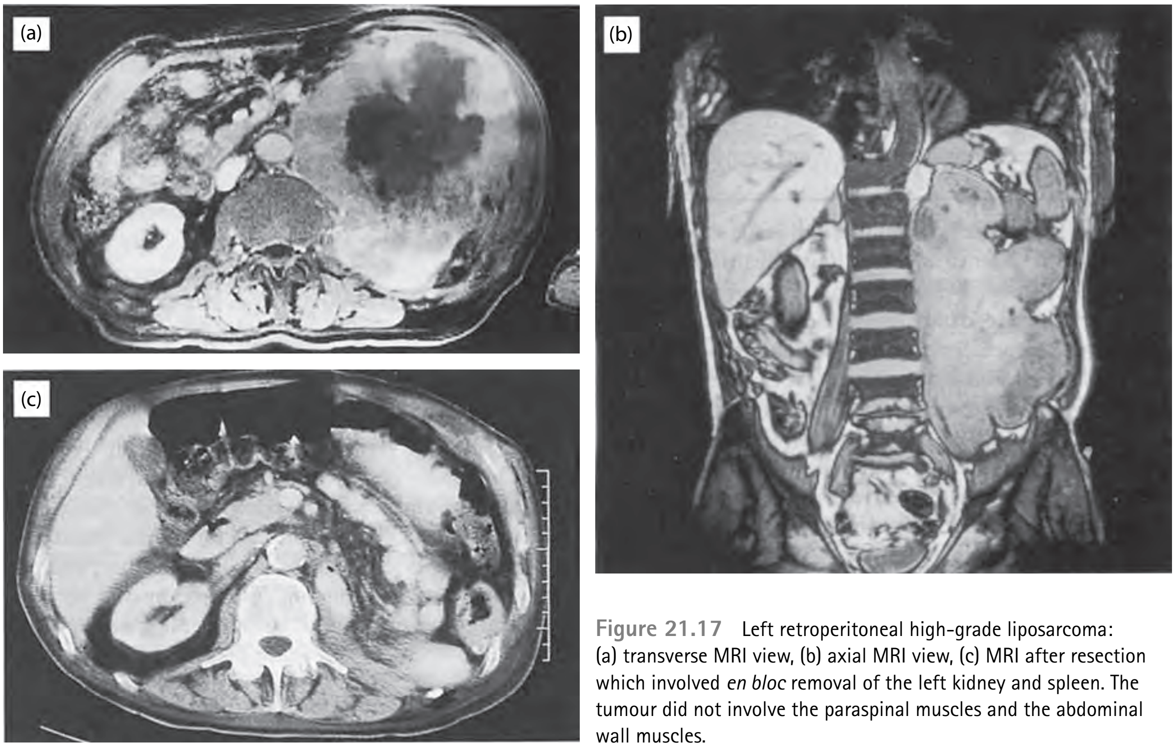


Figure 21.17 Left retroperitoneal high-grade liposarcoma: (a) transverse MRI view, (b) axial MRI view, (c) MRI after resection which involved *en bloc* removal of the left kidney and spleen. The tumour did not involve the paraspinal muscles and the abdominal wall muscles.

Clinical features

Patients with retroperitoneal sarcomas usually present late when the lesion has grown significantly in size, as in the early stages these tumours are totally asymptomatic. The symptoms are non-specific, and include abdominal pain and fullness from pressure symptoms. Other symptoms include back pain, abdominal fullness, anorexia, vomiting fatigue and weight loss. In most instances, however, presentation is with a palpable abdominal mass. If the mass is pelvic and the patient female, there is a real risk of misdiagnosis of gynaecological (usually) ovarian malignancy. This leads to laparotomy and inappropriate transperitoneal biopsy instead of referral to a soft-tissue sarcoma centre. In advanced cases, there may be evidence of caval compression with lower limb oedema, varicocele, ascites and dilated abdominal wall veins. Helical CT is most commonly used for establishing the diagnosis, with intravenous contrast MRI being reserved for the detection of vascular invasion and suspicious small lesions in the liver. Positron emission tomography/18-fludeoxyglucose, if available, is useful for tumour grading (low, intermediate, high) and for the detection of recurrence after excision.

A retroperitoneal core needle biopsy is performed under ultrasound guidance in tumours that appear dedifferentiated on CT. These tests have largely replaced other investigations, although an intravenous urogram is advisable as part of the work-up to exclude ureteric involvement and to assess the function of both kidneys (in case one needs to be removed with the tumour).

Treatment

There is no separate staging system for retroperitoneal sarcomas, although the American Joint Committee Staging System for extremity soft-tissue sarcomas is used in some centres, but its appropriateness for retroperitoneal sarcomas is doubtful. The treatment of retroperitoneal sarcomas, although essentially surgical, is complex and these patients are best managed in tertiary referral soft-tissue sarcoma centres by multidisciplinary teams that have the necessary surgical expertise in dealing with these rare tumours. Distant metastases and infiltration of non-resectable vascular or neural structures preclude resection.

Surgical treatment

The scope of surgical treatment is complete *en bloc* surgical excision with no residual tumour (R0 resection). Although there is agreement on this objective, surgical strategies vary. The orthodox radical *en bloc* resection including macroscopically involved adjacent organs is technically demanding and it often proves difficult even in experienced hands to obtain adequate tumour-free margins. Invasion of the retroperitoneal muscles can pose major problems even in experienced hands. Thus, complete radical resection is possible in only 30% of cases. The resectability rate in published series with the orthodox *en bloc* resection ranges from 38% to 100% but the radical (complete) resection rates are lower.

For these reasons, some centres in France advocate *compartmental resection* on the grounds that an R0 resection can only be achieved by this approach rather then resection *en bloc* with the tumour of macroscopically involved organs. These surgeons report improved survival with the more radical compartmental resection. However, the potential for postoperative morbidity is high and limits the applicability of this approach to all patients with retroperitoneal sarcomas. The problem with compartmental resection is that it entails resection of uninvolved organs to obtain negative margins.

In cases with vascular invasion or in situations when the tumour arises from a major vascular structure such as the inferior vena cava (IVC), combined resection with vascular reconstruction has been described and is safe when performed at a centre with expertise in these complex procedures. Management of the IVC after resection of a leiomyosarcoma is an area of debate regarding the need for reconstruction. Ligation of the IVC without reconstruction is generally well tolerated with a surprisingly low incidence of transient lower extremity oedema or other sequelae, although patients may develop postoperative acute renal failure. In its favour, ligation avoids the long-term complications of grafting, i.e. infection and pulmonary embolism.

Chemotherapy

The role of neoadjuvant chemotherapy is not well defined in view of the rarity of retroperitoneal sarcomas. Neoadjuvant chemotherapy may reduce the extent and complexity of the resection in tumours that respond to systemic therapy, and one retrospective series reported that this is achieved in 13% of patients with a resultant higher resection with negative margins, reduced local recurrence and higher overall survival than non-responders. In other major sarcoma centres preoperative chemotherapy is used selectively for large intermediate- to high-grade lesions with combination doxorubicin/ifosfamide. However, in the absence of data from prospective clinical trials, the reported retrospective case series do not support the universal use of neoadjuvant chemotherapy for all retroperitoneal sarcomas. Instead, neoadjuvant chemotherapy is restricted to biopsy-proven large, intermediate- to high-grade tumours.

Several prospective randomized trials on adjuvant chemotherapy have demonstrated decreased local recurrence rates in patients who received adjuvant regimens. A Cochrane review of 14 trials on adjuvant chemotherapy concluded that local recurrence and disease-free survival were positively affected by adjuvant therapy, but the effect on overall survival was uncertain owing to lack of data. Chemotherapy may also benefit patients with advanced disease in terms of metastasis-free and overall survival.

Radiotherapy

Although commonly used in both neoadjuvant and adjuvant settings, there is no level I evidence for the efficacy of radiotherapy in retroperitoneal sarcomas. However, in most sarcoma centres, preoperative radiotherapy is used in selected cases, usually for intermediate- to high-grade lesions, often as part of neoadjuvant chemoirradiation.

Radiotherapy (external and brachytherapy) is administered in some centres following resection and neoadjuvant radiotherapy.

Prognosis

Prognosis is influenced by four factors: (1) histological grade, (2) completeness of the surgical excision (R0, no residual disease), (3) mitotic activity and (4) multifocal disease. The reported 5 year survival rate after radical resection for cure varies from 62% to 92% in well-differentiated tumours, as distinct from 16–48% in undifferentiated sarcomas. The mitotic activity is a major criterion. The 5 year survival rate is inversely proportional to the mitotic activity of retroperitoneal sarcomas and is worst for sarcomas with more than nine mitoses per 10 HPFs (high risk). The third prognostic criterion is multifocal disease or sarcomatosis (seven or more lesions), which carries a bad prognosis (30% 5 year survival) compared with patients with unifocal disease (60% 5 year survival). The diagnosis of sarcomatosis is made when detailed examination reveals seven or more lesions.

Hernias

Internal hernias

Internal hernias are a rare but important cause of intestinal obstruction (0.2–0.9% of all cases) because they are often undiagnosed before emergency laparotomy; not uncommonly, they lead to infarction necessitating bowel resection (of varying extent), and this contributes to the high morbidity and mortality rates. Internal hernias are often classified as developmental (congenital) or acquired. In the congenital types, except in instances of gross mid-gut malrotation, the presentation occurs over a wide age range, but the majority of patients are in the fifth decade. By definition, developmental internal hernias cause obstructive symptoms in the absence of any previous surgical intervention.

Acquired internal hernias occur in patients who have undergone abdominal surgery, most commonly of the upper gastrointestinal tract, e.g. after gastric bypass surgery for morbid obesity or following Roux-en-Y diversion (especially if this is antecolic). They are well documented after right hemicolectomy (both open and laparoscopic), after laparoscopic extraperitoneal hernia repair (through unrecognized defects in the peritoneum), vascular operations, e.g. intraperitoneal femorofemoral bypass grafting, and after pelvic lymphadenectomy when loops of small intestinal become trapped beneath one or other iliac arteries. A particular group of acquired internal hernias occurs, albeit rarely, in relation to transplantation of the kidneys and liver. These are sometimes referred to as paratransplantation internal hernias and can have serious consequences. Following renal transplantation the internal hernia is caused by entrapment of bowel or omentum through a defect in the peritoneum covering the transplanted kidney. Internal herniation with volvulus of the small intestine is a potentially fatal complication after liver transplantation. The herniation occurs around the Roux-en-Y loop used for the biliary reconstruction. The mortality of this condition (from graft or bowel necrosis) is 50%. In view of the high mortality of peritonitis in transplant patients, early surgical treatment is indicated in all these patients who develop intestinal obstruction after surgery.

Pathological anatomy

This provides a more useful classification and better understanding of the underlying pathology. Thus, internal hernias fall into two groups: *true hernias* and *internal prolapses.*

True hernias

These have a hernial sac. The orifice involved may be:

- normal, e.g. Winslow's foramen
- paranormal, i.e. peritoneal fossae – paraduodenal, ileocaecal, inter- and mesosigmoidal, paracolic, supravesical.

Internal prolapse

These do not have a sac. The prolapse may occur through:

- an abnormal orifice in a mesentery or an omentum (transmesenteric, transmesocolonic, transomental)
- an anomalous orifice – congenital defects in a ligament (falciform, gastrosplenic, broad ligaments) or a mesentery (mesentery of Meckel's diverticulum)
- surgical defects or entrapment by surgically altered internal anatomy, e.g. anastomoses, stomas.

Paraduodenal hernias (right and left in close relation to the fourth part) constitute the commonest true internal hernias and account for 30–53% of cases. They are caused by incomplete rotation of the mid-gut such that the small intestine becomes entrapped behind the colon and mesocolon. The most valuable investigation for paraduodenal hernia is a small bowel enema, which usually shows clumping of the small intestine with incomplete rotation of the caecum and ascending colon. Other true internal hernias worthy of specific mention are hernia through the foramen of Winslow (epiploic foramen) and paracaecal hernia. Herniation through the epiploic foramen accounts for 8% of internal hernias. Most commonly the caecum and ascending colon are involved, and, in some cases, re-enter the main peritoneal cavity through an additional congenital defect of the lesser omentum. In patients with chronic symptoms, a barium enema may be diagnostic because it shows the caecum lying posterior and medial to the stomach. Paracaecal hernia presents with both subacute and acute low small bowel obstruction.

Internal hernia (prolapse) arises as a complication of gastrointestinal surgery in 0.3–1.0% of cases. The majority occur after upper gastrointestinal surgery (Billroth II gastrectomy with antecolic gastrojejunostomy and enteroanastomosis, Roux-en-Y reconstruction or diversion, etc.) and less commonly after colon surgery (right hemicolectomy, colostomy). Presentation with intestinal obstruction may occur early or several years after (up to 25) the intervention.

Malrotation of the mid-gut usually presents acutely in the neonatal period but affected individuals may not develop symptoms until much later (2–23 years). The most common symptoms include vomiting (68%), colicky abdominal pain (55%) and diarrhoea (9%). The diagnosis is made by barium meal and follow-through, or, preferably, a small bowel enema. A significant number of these late presenting patients (40%) are found to have either a volvulus or internal hernia at operation.

Clinical features

The reported incidence of all internal hernias varies between 0.2% and 0.9% of autopsies, 0.3–2.0% of parietal hernias and 0.01% of laparotomies. Overall the condition is more common in males (male to female ratio 3:2). The age distribution varies widely but peak symptomatic incidence is in the fifth decade (mean age of 45–50 years). Some (10–15%) are discovered as an incidental finding during laparotomy for another condition. Most commonly, these are paraduodenal hernias.

The symptoms of internal hernia irrespective of type are entirely non-specific and very few cases are diagnosed in the elective situation, with the vast majority (90%) presenting with acute intestinal obstruction, which is often strangulating with evidence of established peritonitis caused by infarcted bowel (30–60%). In a minority of patients recurrent attacks of colicky pain and abdominal distension are followed by imaging tests (contrast radiology, CT scanning, abdominal ultrasound), and these may provide a preoperative diagnosis. More rarely still, a mass is found on physical examination of the abdomen in a patient with subacute obstructive symptoms. Acute intestinal obstruction with strangulation occurring in a patient without an external hernia or previous abdominal surgical intervention should suggest the possibility of a congenital internal hernia, especially if the patient gives a history of chronic intermittent abdominal pain and a palpable abdominal mass is found on examination.

Treatment

The essence of good management is early intervention (laparotomy) as this is the only means of preventing necrosis of the bowel. The surgical treatment consists of reduction of the hernial contents, resection of the necrosed bowel, usually with primary anastomosis, and correction of the anatomical defect that caused the herniation in the first instance. The operative reduction of the hernia and the surgical repair must be conducted with extreme care to avoid injury to the major mesenteric vessels juxtaposed to or surrounding the hernial orifice. The hospital stay, morbidity and mortality (up to 30% in large series) depend on the presence/absence of bowel infarction. In the rare instances when an internal hernia is discovered after investigation of chronic symptoms, elective surgery is needed because of the pathogenic potential of this condition.

Diaphragmatic hernias

Strictly speaking diaphragmatic hernias should include hiatal hernias but, by convention, these are included with gastro-oesophageal reflux disease (see Chapter 22). Diaphragmatic hernias excluding this category fall into three groups:

1. congenital diaphragmatic hernia
2. traumatic diaphragmatic hernia – secondary to undiagnosed rupture of the diaphragm
3. herniations through congenital small defects – through the foramen of Morgagni (anterior) and foramen of Bochdalek (posterior).

Congenital diaphragmatic hernia

This is a serious congenital anomaly, which is always associated with pulmonary hypoplasia and other abnormalities (35%), particularly of the central nervous system and the skeleton. The condition is diagnosed in the majority (80%) by prenatal ultrasound, when the question of termination arises. With emergency neonatal surgery (reconstruction of the diaphragm with prosthetic mesh) and extracorporeal membrane oxygenation for cases with severe pulmonary impairment, the survival of infants born alive with congenital diaphragmatic hernia averages only 60% and more than half (60%) of these have persistent disorders that include respiratory problems, developmental delay, poor growth and gastro-oesophageal reflux. Despite advances in neonatology there is still a high mortality and morbidity associated with congenital diaphragmatic hernia.

Rupture of the diaphragm and traumatic diaphragmatic hernia

The primary event is rupture of the diaphragm owing to severe blunt trauma, most commonly from motor vehicle accidents. The rupture, which is twice as common on the left side and is rarely bilateral (5%), may be diagnosed immediately or missed when a traumatic hernia is diagnosed because of symptoms or complications. Rupture of the diaphragm is more common in males (male to female ratio 4:1) and occurs over a wide age range. Associated injuries are present in the majority of patients (90%) with rupture of the diaphragm. These include rib fractures, splenic, hepatic, pulmonary and bowel injuries. Helical CT with sagittal and coronal reformatted images is the best imaging test for the diagnosis of diaphragmatic rupture after blunt trauma. Findings consistent with diaphragmatic injury include waist-like constriction of abdominal viscera (collar sign), intrathoracic herniation of abdominal contents and diaphragmatic discontinuity. Helical CT reformatted images detect 78% of left-sided and 50% of right-sided diaphragmatic injuries.

The injury is always a major one with a reported mean injury severity score (ISS) of 30–35. The immediate mortality ranges from 15% to 20%. Predictors of death include old age, high ISS, severe hypovolaemic shock and bilateral injuries. Emergency repair of a ruptured diaphragm can be carried out through a laparotomy, thoracotomy or the thoracoabdominal approach, depending on the nature of the associated injuries.

Traumatic diaphragmatic hernia may be diagnosed months to several years (average 5 years) after the injury. The patient may have non-specific symptoms (vague chest pain, shortness of breath, palpitations) or presents acutely with clinical signs of intestinal obstruction/strangulation of a hollow organ. In line with its aetiology traumatic diaphragmatic hernia is more common on the left. The chest radiograph is often suggestive and diagnosis is usually confirmed by barium studies of the gastrointestinal tract. The most common herniated abdominal organs in traumatic diaphragmatic hernias are the stomach and colon. The repair involves closure of the defect by a prosthetic mesh and this can be performed through a thoracotomy or laparotomy or laparoscopically.

Iatrogenic diaphragmatic hernia

This is due to damage to the diaphragm during surgery, particularly laparoscopic antireflux surgery. The injury is often caused by high-frequency electrosurgery and the resulting defect, which is small, leads to herniation, usually of the stomach over the succeeding months. The most common complaint is pain in the left upper quadrant and left shoulder. The condition is diagnosed by a barium meal and is often missed by flexible endoscopy.

Morgagni and Bochdalek hernias

Hernias through the foramina of Morgagni (right and left) are rare congenital hernias that may present in infancy, childhood and adult life. They are situated anteriorly in the immediate retrosternal position. In children Morgagni hernias are usually asymptomatic but they may be associated with mild respiratory distress or cause gastrointestinal symptoms and rarely incarceration of bowel, which may be of the partial Richter type. In adults, they may present with vague symptoms or intestinal obstruction or gastric volvulus. The foramina of Bochdalek are congenital posterior diaphragmatic defects resulting from persistence of the pleuroperitoneal canal on one or both sides. Bochdalek hernias are rare and present in adult and elderly patients with digestive symptoms and less frequently with obstruction. Both hernias do not have a sac and are thus instances of internal prolapses.

Incisional hernias

Incisional hernias after abdominal surgery are common but with a varied reported incidence depending on case mix (2–20%). Strictly speaking, the term ventral hernia should be restricted to incisional hernia arising in abdominal midline operative wounds, but often the two names are used synonymously. There are several factors that contribute to the aetiology of incisional hernias, the most important being adequacy of abdominal wound closure in the first instance and the occurrence of wound infection and subclinical wound dehiscence in the postoperative period. There is no evidence that abdominal wound closure with synthetic monofilament biodegradable sutures carries a higher incidence of incisional hernia than closure with non-absorbable sutures. There are however a number of recognized risk factors, which include:

- wound infection: most important aetiological factor
- postoperative abdominal distension: pressure on wound
- obesity, especially morbid obesity: very strong risk factor
- chronic obstructive airways disease
- emergency surgery: independent risk factor
- reoperation for a postoperative complication: independent risk factor
- type of incision: more frequent after vertical than transverse/oblique incisions
- steroid dependence: less important than morbid obesity
- underlying disease: diabetes
- creation of a stoma (colostomy, ileostomy, urinary conduit): parastomal hernias
- age >70 years
- no prophylactic antibiotic cover during primary operation: increased incidence.

Wound infection is the most significant independent factor for incisional ventral hernia. In patients who develop postoperative wound infection, the reported risk of incisional hernia averages 25%. Many reports indicate that transverse/oblique incisions are accompanied by a lower rate of incisional hernias than vertical midline incisions. There is some evidence that, apart from technical faults and risk factors, incisional hernia may arise in patients with an underlying collagen metabolic disorder, known to play an important role in the development of inguinal hernia. Studies have shown a decreased ratio of collagen I/III owing to a concomitant increase in collagen III in patients with incisional and recurrent incisional hernias. This appears to reduce the mechanical strength of connective tissues and may explain the high incidence of recurrence in patients undergoing fascial repair procedures.

Clinical features

Aside from the obvious disfigurement, they cause pain and discomfort and may strangulate, although the risk of this complication is low. The diagnosis is obvious and is made on clinical examination, when tensing of the muscles of the anterior abdominal wall (elevation of head from pillow or lower limbs from bed) accentuates the bulge. The vast majority of incisional hernias are reducible when the patient lies down in the supine position.

Treatment

This has to be individualized. Thus, if the patient has minimal symptoms and especially if elderly or suffers from significant comorbid disease, conservative management with an abdominal support is advisable. Many patients, however, have symptoms and dislike the bulge, and for this reason insist on repair. There are many techniques of surgical repair of incisional hernias:

- primary repair
- primary repair with relaxing incisions
- primary repair with onlay mesh reinforcement
- onlay mesh only
- inlay mesh placement
- retrorectus mesh placement
- intraperitoneal mesh placement.

There is now sufficient reported evidence that primary fascial repair techniques have an unacceptably high incidence of recurrence (up to 50%) and these operations have largely been abandoned in most centres. By contrast, tension-free repair by a synthetic mesh is accompanied by a much lower recurrence rate (2–10%) and is the technique that is in general use.

Mesh materials

Absorbable meshes are only used in cases when mesh infection is a significant risk and primary closure is not possible. Polyester mesh appears in some studies to be associated with higher rates of mesh infection and enterocutaneous fistula formation. However, fluorinated polyester mesh that can be gel impregnated for antibiotic bonding immediately before use has been shown experimentally to induce minimal adherence to bowel and this material exhibits minimal contraction and hardening with time.

Polypropylene has greatest tissue ingrowth of all non-absorbable mesh materials. Polytetrafluoroethylene (PTFE) (Gore-Tex) has the lowest bowel adherence rate and related complications. The Gore Dualmesh Plus, which incorporates antimicrobial agents (silver and chlorhexidine), carries the lowest mesh infection rates in most reported studies but is expensive. One experimental study involving infection of graft materials with *S. aureus* and meticillin-resistant *S. aureus* (MRSA) organisms showed that Gore Dualmesh Plus biomaterial was the only mesh material used in the study able to kill both *S. aureus* and MRSA.

The most commonly used are polypropylene meshes, which may be monofilament (Marlex), double filament (Prolene) or multifilament (Surgipro, Atrium, etc.). The absorbable meshes available are polyglactin 910 (Vicryl) and polyglycolic acid (Dexon). Whatever material is used, it is important that the mesh overlaps the size of the defect by a significant margin, and it should lie loosely rather than be stretched over the defect. Tension-free mesh repair can be undertaken laparoscopically with good results.

The types of repair techniques for incisional ventral hernia are:

- *Primary repair*: this consists in freshening and approximating the two edges of the defect. It is no longer practised in view of high recurrence (>50%) owing to tension on the repair. Some procedures (Keel repair) use relaxing incisions to reduce tension but still incur an appreciable recurrence.
- *Primary closure with mesh reinforcement*: an onlay is sutured over the primary repair to the anterior rectus sheath as reinforcement. However, the hernia repair suture line remains under tension. In addition, the mesh increases the infection risk without any material benefit in terms of recurrence.
- *Inlay mesh repair*: after excision of the hernial sac, the fascial margins of the defect are sutured to a mesh [polypropylene or expanded PTFE (ePTFE)]. Polypropylene is used when omentum can be placed between intestine and mesh; otherwise, ePTFE is preferred. This method is difficult to perform laparoscopically and is usually carried out by the open approach.
- *Retrorectus mesh repair (Rives–Stoppa technique)*: in this technique the hernia sac is used to separate the mesh from the intra-abdominal contents. Above the umbilicus, the dissection to create the necessary space for the mesh is performed above the posterior rectus fascia and under the rectus muscles; below the umbilicus, dissection is in the preperitoneal space. A suitably sized mesh is placed in the dissected place and fixed to the muscle layer superficial to the mesh. This repair is safe and effective with a reported low recurrence and morbidity rates.
- *Intraperitoneal underlay mesh repair*: this technique can be performed by either the open or laparoscopic approach, which is increasingly favoured nowadays.

Complications of tension-free mesh repairs

The risks of tension-free repair of incisional hernias include:

- wound infection
- infection of the mesh
- seroma formation
- wound sinus
- enterocutaneous fistula formation
- recurrence.

Although wound infection is common (4–17%), actual incidence rates of mesh infection are not known but most such instances resolve with antibiotic treatment, although persistent serious infections necessitating removal of the mesh are well documented. This is a major problem. The management is staged with temporary skin cover until infection is eradicated, followed by abdominal wall reconstruction by muscle transfer flaps. Seroma formation is also common but its incidence varies (0–22%) with the size of the defect and hence the mesh. Most surgeons insert tunnelled Redivac suction drains to the site to prevent this complication. There is no evidence that these tunnelled small-calibre suction drains increase the incidence of postoperative wound/mesh infection. However, the efficacy of these suction drains in reducing the incidence of seroma formation remains unproven. Wound sinus is reported in 4–18% of cases. It usually heals with conservative management or minor intervention to remove the offending suture. Enterocutaneous fistula has been reported but appears to be a rare complication. Its exact aetiology is not known but adherence of the bowel to the mesh appears to initiate the process. In a large retrospective report on 136 cases, the patient-related factors considered important on statistical evaluation in relation to recurrence following tension-free mesh repair with polypropylene were:

- age >70 years
- hernia >6 cm
- no prophylactic antibiotic cover
- recurrent hernia
- wound infection.

It is important to stress that the recurrent hernia is usually larger than the initial one and hernial defects >10 cm may require autogenous reconstruction of the abdominal wall by muscle flaps.

Massive midline wall defects

These include large central incisional hernias (usually recurrent) and pose major problems in management. In one reported series of 22 patients, the defects varied in size from 14 to 24 cm and the causes included:

- removal of infected synthetic mesh material (32%)
- recurrent incisional hernia (18%)
- removal of split-thickness skin graft and dense abdominal wall cicatrix (18%)
- parastomal hernia (9%)
- primary incisional hernia (9%)
- trauma/enteric/abdominal wall sepsis (9%)
- abdominal wall tumour resection (4.5%).

In view of the size of the defect, conservative management with an abdominal support is unsatisfactory, and, if the patient is fit, surgical treatment is indicated. Reconstruction of these large central abdominal wall defects is a major surgical undertaking that requires specialist plastic surgical expertise. Modern surgical treatment is based on autogenous tissue reconstruction introduced by Ramirez *et al.* The technique utilizes bilateral, innervated, bipedicle, rectus abdominis–transversus abdominis–internal oblique muscle flaps that are transposed medially to reconstruct the central defect.

The results of the Ramirez autogenous tissue reconstruction are good, with satisfactory healing and no recurrence in 90%. The morbidity considering the magnitude of the operation is reported to be minimal and confined to superficial infection and wound seroma formation. The postoperative mortality averages 4–5%. Although the Ramirez operation is usually used to treat complicated (trauma, surgical excision) or recurrent central abdominal wall defects, it is also used as a primary repair of large central incisional hernias.

Port site wound (incisional) hernias after laparoscopic surgery

The overall reported incidence of incisional hernias through port site wounds is 2–3%. The risk factors include:

- size of port >10 mm
- obesity
- inadequate fascial closure.

The vast majority occur in port wounds >10 mm but instances of herniation have been reported in smaller port wounds (5 mm), especially in the lower abdomen. It has to be remembered that, below the arcuate line, there is no posterior rectus sheath and thus the abdominal wall is intrinsically weaker. The commonest reported site is the umbilical region, probably because this is usually a large port (used to insert the laparoscope and extract specimens). Patients can present with localized pain and a subcutaneous lump, which is tender and has a cough impulse; or acutely with acute intestinal obstruction. This usually involves the small bowel and is of the Richter type. In obese patients, the small hernia can be easily overlooked as a cause of the small bowel obstruction (see below).

Parastomal hernias

A parastomal hernia is an incisional hernia that occurs at the site of a surgically constructed intestinal stoma on the abdominal wall. The basic underlying cause for the development of a parastomal hernia is progressive enlargement of the trephine opening in the abdominal wall, owing to tangential abdominal forces working on the circumference of the opening through which the bowel emerges. This physical consideration accounts for the unsatisfactory results of surgical repair irrespective of its nature.

Parastomal hernias are common and their management is both difficult and controversial. They continue to pose management problems. In one large series of 316 patients with 322 stomas, an overall incidence of 67% stoma-related complications was reported. This included a 31% parastomal herniation rate, stenosis of the stoma in 10% and prolapse in 7% at 10 years after surgery. Parastomal hernias include:

- paracolostomy hernia
- ileostomy hernia
- urinary conduit hernia.

After sigmoid colostomies the crude and actuarial risk of complications are 50% and 58%, respectively, at 13 years with paracolostomy hernia being the highest at 35–40% at 10 years. Paracolostomy hernias are more likely in the elderly and in patients with other abdominal wall hernias. From the published data, the extraperitoneal technique appears to reduce the incidence of paracolostomy hernia. Other technical factors, e.g. mesenteric fixation and siting the stoma through the belly of the rectus abdominis, do not appear to influence the rate of this complication, although there are strong proponents for both measures.

In general, complications are detected much later in patients with a urological stoma than in those with a colostomy. The high incidence of stomal complications requires long-term follow-up of these patients.

Paracolostomy hernia most commonly presents with a bulge, ill-fitting bag and leakage problems, but they may present acutely with intestinal obstruction (usually small bowel), and indeed strangulation. Instances of incarcerated stomach in the hernial sac have been reported. The presumed reduction in the risk of intestinal obstruction with closure of the lateral space has not been confirmed by long-term studies. However, there is a slight reduction in the risk of intestinal obstruction with the extraperitoneal technique.

The stomal complications of ileostomy may occur many years after construction and at 20 years the incidence of stomal complications exceeds 70% in patients operated on for ulcerative colitis, but is lower – though still high (50–60%) – in patients following colectomy for Crohn' disease. The complications in order of frequency are:

- skin problems (34%)
- intestinal problems (23%)
- retraction (17%)
- parastomal herniation (16%)
- prolapse.

Thus, stomal herniation is distinctly less common after terminal ileostomy than after terminal colostomy. Again closure of the lateral space does not diminish the risk of intestinal obstruction and fixation of the mesentery does not reduce the probability of prolapse of the ileostomy. The incidence of parastomal herniation is unaffected by siting the ileostomy through the rectus abdominis as distinct from the oblique muscles.

The incidence of parastomal hernia in patients with ileal conduit diversion is lower than ileostomy after proctocolectomy (4–5%). The most common presentation is with a poorly fitting appliance causing leakage of urine. However, acute presentation with obstruction, anuria and parastomal ileal conduit fistula are well documented.

Treatment of parastomal hernias

The surgical treatment of parastomal hernias presents a continuing challenge and there is no universally effective operation. There are four techniques used in the surgical treatment of parastomal hernias:

1. relocation of the ostomy with repair of the defect (fascial or with mesh)

2 mesh repair of the defect around the ostomy exit site; in this method, the hernia sac (laparocoele) is replaced without being opened and the mesh positioned in the preperitoneal space
3 reduction and mesh repair with two strips of polypropylene mesh through a midline incision
4 special prosthesis consisting of a polypropylene ring mounted in the centre of a polypropylene mesh.

Irrespective of technique, parastomal hernia repair is often unsuccessful (average recurrence rate of 30% at 5 years) and rarely without complication. The general consensus for first-time repair is to relocate the stoma and repair the defect with a tension-free mesh. This, of course, may transfer the problem to the other side. For recurrent parastomal hernias, local repair with prosthetic material (without relocation) is advocated, as it appears to be the best of a bunch of poor alternatives. In either case, fascial repair alone should not be performed owing to an unacceptably high recurrence rate.

There have been other techniques reported but the data on their efficacy are limited. As the basic problem is progressive enlargement with time of the hole by the tangential abdominal forces, a ringed prosthesis technique has been reported, which appears to give good results in the short term. The prosthesis consists of a polypropylene ring of varying internal diameter (20, 25 or 30 mm), mounted in the centre of a polypropylene mesh. Following mobilization the exteriorized bowel is inserted through the ring and the mesh sutured to the parieties. In a series of 14 patients treated with this technique there was only one recurrence during a follow-up period of 5–35 months. Another procedure uses the midline laparotomy approach and, after reduction, two strips of polypropylene are sutured on either side of the bowel to prevent enlargement of the orifice.

Non-incisional abdominal wall hernias

These hernias occupy a good deal of surgical time and account for 10–15% of all surgical operations. The majority of operations (80%) are performed for inguinal hernias, although this figure is even higher in the male population. The remainder are in the region of the umbilicus (8%), incisional (7%) and femoral hernias (5%). Rarer forms of hernias, although very interesting, form only a tiny proportion of the surgical problem.

Diagnosis of abdominal wall hernias

In the vast majority of patients, the diagnosis is made on history and physical examination (location of the bulge and cough impulse) and no other confirmatory tests are needed. However, diagnostic problems may be encountered especially in obese patients and those presenting with acute intestinal obstruction. Water-soluble contrast herniography is an accurate means of identifying inguinal and femoral hernias in cases presenting diagnostic problems. For other hernias, ultrasound and cross-sectional imaging CT or MRI is preferred. These are especially useful in the detection of the rare, e.g. Spigelian or obturator (pelvic floor), hernias.

Epigastric, umbilical and paraumbilical hernias

These hernias are grouped together because herniation occurs through a defect in the linea alba between the xiphisternum and the umbilicus. The paraumbilical hernia is an epigastric hernia situated just above the umbilicus.

Epigastric hernia

This hernia is usually encountered in males above the age of 40 years. About one-quarter of cases are multiple. The defect in the linea alba allows a small pad of extraperitoneal fat to protrude, and, as a result of the increased intra-abdominal pressure, the defect enlarges and then permits a sac of peritoneum, and eventually the sac may admit omentum or even small bowel. Most epigastric hernias are symptomless and are diagnosed incidentally by the patient or doctor. A small nodule, which is more prominent on standing, is palpable in the midline but it is rare for a cough impulse to be elicited. A small number of patients present with vague upper abdominal symptoms that do not fit a dyspeptic pattern and in whom repair of the hernia gives relief. It may be that in these patients a degree of tension on the peritoneal sac has produced the symptoms.

Small epigastric hernias do not require treatment. Those >2.0 cm have the potential of strangulation because of the narrow neck through the linea alba and are best treated surgically. The procedure consists of excision of the sac and repair of the defect by either a simple longitudinal fascial repair or a transverse overlapping of the Mayo type. The results except in the very obese are generally very good.

Adult paraumbilical hernia

These hernias are more common in females (female to male ratio 3:1) and are largely confined to obese patients. The defect lies just above the umbilicus, although deformity of the umbilical button is the earliest manifestation. In adults these are common with an overwhelming female preponderance. In one study of 2100 patients undergoing laparoscopy, paraumbilical fascial defects were found in 18% of patients and only 56% of these had symptomatic or overt umbilical hernias. The hernia may enlarge to the size of an orange but the neck of the sac remains dangerously small so that the risk of strangulation is ever present. The chronicity of the condition leads to firm adhesions forming between the peritoneal sac and its contents so that almost all large hernias are irreducible and should never be treated by the use of an abdominal support or truss. Surgical treatment is best carried out effectively in all patients. The majority of patients, however, present acutely with obstruction/strangulation. The procedure of choice for repair of paraumbilical hernia is that described by W.J. Mayo in 1893. After excision of the sac, the defect is closed by suture, the upper crescent being fixed over the front of the lower half of the defect. The Mayo operation has reported recurrence rates of 2–3%. For large hernias in the elective situation, some now use a prosthetic mesh repair.

Infantile umbilical hernia

The worst defect is exomphalos, which fortunately is a rare condition, occurring in about one in 5000 births and is nowadays diagnosed prenatally. It is frequently associated with other congenital defects and not surprisingly about one-quarter of the babies have malrotation of the intestine.

Simple umbilical hernia is common in infants and young children with the highest reported incidence in African Caribbean babies. In the vast majority of infantile umbilical hernias spontaneous closure occurs (reduction in size by approximately 18% each month) before the age of 4 years and thus management is conservative, especially as complications during this age period are rare. However, large infantile umbilical hernias (neck of the sac >2 cm) are unlikely to close spontaneously and surgical repair is therefore advisable. Irrespective of size, an umbilical hernia that persists beyond 4 years requires surgical repair, as closure then becomes unlikely. It is generally considered important to preserve the umbilical cicatrix after excising the sac and repairing the defect so that the child will not appear different from its fellows.

The most common complication in these children is incarceration with the development of small bowel obstruction. The reported risk of this complication is approximately one in 1500 umbilical hernias. Another very rare complication reported in Nigerian children is spontaneous rupture. This usually occurs in the first year of life and is probably precipitated by raised intra-abdominal pressure from excessive crying. The condition results in partial evisceration and needs urgent intervention.

Adult umbilical hernia

The umbilical hernia that develops in patients with refractory ascites due to chronic liver disease is often overlooked but can assume clinical significance. In the first instance incarceration is well documented as a complication of effective relief of the ascites following diuresis, paracentesis, peritoneovenous shunting and transjugular intrahepatic portosystemic shunt. Second, these patients have marked atrophy of the abdominal muscles and some have, in addition, large high-pressure veins following recanalization of the umbilical vein (caput medusae). Thus repair can be difficult and should always be conducted using prosthetic mesh, preferably of the double-stranded, closely knitted variety. In addition, subcutaneous suction drains should be avoided because of the increased risk of infection of the ascitic fluid. The use of prophylactic antibiotics against Gram-negative bacteria is mandatory.

Groin hernias

Because of the erect posture, the inguinofemoral area is subjected to maximum strain from the intra-abdominal pressure, especially during exertion. This together with the intrinsic weakness that is caused by the inguinal canal in males accounts for the predominance of these hernias in males. The hernia starts as a small pressure-induced diverticulum (hernial sac) that emerges through the deep inguinal and less commonly the femoral ring to enter the respective canal and exits as a lump in the groin that has a cough impulse. The term groin hernia covers inguinal and femoral hernias, which are discussed separately.

Complications of groin hernia

The complications of groin hernias are:

- irreducibility
- obstruction
- strangulation.

The majority of patients who are admitted as emergencies with complicated groin hernias have not previously sought medical attention or been diagnosed with the condition in the outpatient department. This observation implies that most hernias that develop complications do so within a relatively short time in the natural history of the disease. Mortality of obstructed hernias is high in patients with coexisting cardiorespiratory disease, whereas the morbidity rate is influenced by the viability of contents of the hernial sac. In turn, this is directly related to the duration of irreducibility/incarceration or delay in presentation.

The risk factors for complications of groin hernias are:

- Adults:
  - old age
  - duration of hernia: short duration <1 year
  - type of hernia: femoral more than inguinal
  - coexisting medical illness: especially chronic obstructive airways disease.
- Children:
  - very young (infants)
  - gender (male)
  - short duration of hernia
  - side (right).

Inguinal hernia

By far and away, inguinal hernia is the most common external abdominal hernia, accounting for over 90%. Normally, two mechanisms act to prevent herniation through the inguinal canal. With increased intra-abdominal pressure, contraction of the internal oblique and transversus muscles acts upon the section of the transversus aponeurosis that arches convexly upwards over the medial half of the canal. The arch is pulled down and flattened and thus reinforces the posterior inguinal wall. The second mechanism, which may in fact be more important, depends upon the attachment of the strong fascial layer forming the deep inguinal ring. This fascial ring is normally firmly adherent to the posterior surface of the transversus muscle so that contraction of this muscle pulls the ring upwards and laterally.

In essence, an inguinal hernia is the consequence of weakness of the posterior abdominal wall. In the past, stretching of the transversalis fascia was considered to be the most important factor and some fascial repair procedures (Shouldice operation) were based on suture plication of this layer. Some, however, consider the transversalis (endoabdominal) fascia to be the thinnest and least important layer in the prevention of inguinal hernia formation and consider that the strength of the posterior wall of the inguinal canal is due to the muscle fibres and aponeuroses of the internal oblique and transversus abdominis. There is evidence for an increased risk of right inguinal hernia after appendicectomy. This is related to the denervation of the right transversus abdominis muscle fibres. These fibres are responsible for the support of the deep inguinal ring of fascia. Weak fascial support due to an abnormal accumulation of type III collagen has been demonstrated in some patients with direct but not indirect inguinal hernia.

Inguinal hernia exhibits a marked male predominance (male to female ratio 20:1) and from the anatomical standpoint is of two types – indirect (or lateral) and direct (or medial). The sac of an indirect hernia arises from the processus vaginalis. The hernia travels down the inguinal canal from the internal ring and exits as a subcutaneous lump, exhibiting a cough impulse at the external inguinal ring above and medial to the pubic tubercle. The herniation lies inside the spermatic cord (covered by all three spermatic fascial layers). In view of its indirect course, it does not often reduce itself spontaneously when the patient lies down and is more prone to irreducibility than the direct inguinal variety. Indirect inguinal hernia enters the scrotum as it enlarges, whence it qualifies as inguinoscrotal hernia (Figure 21.18). The same pathology occurs in the female.

Persistence of the processus vaginalis forms a peritoneal diverticulum (canal of Nuck). This enters the inguinal canal to form the hernia, which, after exiting from the superficial inguinal ring, enters the labium majus as it enlarges. The direct hernia results from a weakness of the posterior wall of the inguinal canal medial to the internal ring and hence the inferior epigastric vessels; the sac is thus in close proximity to the external ring. Occasionally, the medial wall of the sac of a direct inguinal hernia is composed of the bladder wall. As the hernia enlarges it exits through the superficial inguinal ring behind or above the spermatic cord. Because of the direct path and wide neck, the vast majority of direct hernias are spontaneously reducible when the patient lies down and rarely strangulate. Moreover, even when large, direct hernias do not enter the scrotum; they tend to enlarge in the groin beneath Scarpa's fascia, which obstructs their entry into the scrotum. Patients with major weakness of the posterior inguinal wall may develop combined direct and indirect hernias with the two hernial sacs straddling the inferior epigastric vessels – this is sometimes referred to as a pantaloon hernia.

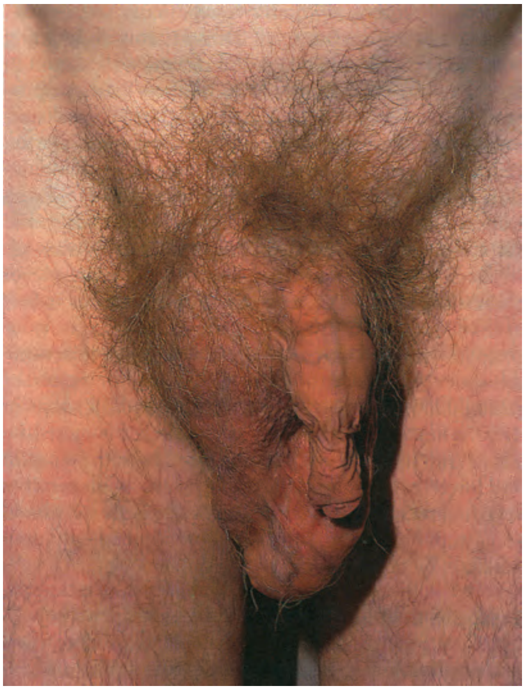


Figure 21.18 Inguinoscrotal hernia.

Clinical features

The complaint common to nearly all patients is the appearance of a lump in the groin. Some patients complain of a dragging sensation or pain in the groin, particularly during the early stages, but many hernias are asymptomatic. There may be a history of a major physical strain or of heavy physical work prior to the development of the hernia. Presumably, increased abdominal pressure can stretch the fascial margins of the deep inguinal ring and open up a preformed peritoneal sac.

The hernia may be best shown with the patient standing or precipitated by coughing and straining. The bulge should be above and medial to the pubic tubercle. In patients with large inguinoscrotal hernias the diagnosis is obvious though a cough impulse may be difficult to elicit. The differential diagnosis of inguinal hernia includes femoral hernia, inguinal lymphadenopathy, ectopic testis, hydrocele of the cord, saphena varix and lipoma of the cord. Endometriosis in a hernial sac may be mistaken for an incarcerated hernia, as may malignant tumours in the groin. Patients with chronic groin pain constitute a difficult diagnostic problem. Most cases of chronic pain in athletes are due to soft-tissue injuries, but, in a small number, an undeclared hernia is the cause.

In the vast majority of cases, the diagnosis of an inguinal hernia is made clinically and a good physical examination can distinguish between the two types of hernia. The old practice of inserting a finger into the inguinal canal to detect the exact position of the cough impulse (tip of the finger in indirect and volar aspect of the finger in direct hernia) is no longer practised because it causes considerable patient discomfort. The best technique is to reduce the hernia with the patient in the supine position, and then apply pressure over the internal ring. This should control the hernia only if it is of the indirect variety. When difficulty arises in establishing the diagnosis because the hernia is small, especially if the patient is obese, water-soluble contrast herniography (50–80 mL injected intraperitoneally) may be used.

Sliding inguinal hernia

Essentially, in a sliding inguinal hernia, part of the sac wall is formed by prolapsed viscus and, for this reason, these hernias are always large (Figure 21.19). The condition can be suspected but not confirmed preoperatively on clinical examination. Usually the diagnosis is established during surgery, although ultrasound and cross-sectional imaging may identify sliding components preoperatively. The most common organ associated with a sliding inguinal hernia is the sigmoid colon, but other organs can be involved. These include the vermiform appendix, urinary bladder, etc. In neonates, large sliding hernias containing Fallopian tubes, ovaries and uterus are well documented and such instances have been reported in middle-aged women.

The more usual sliding hernia involving the colon forms a large inguinal mass and is often irreducible by the time of diagnosis. Sliding inguinal hernia may present acutely with strangulation involving omentum, small bowel, bladder or colon itself.

Inguinal hernias in children

In this age group, 90% of inguinal hernias occur in males, and these often present at about 1 year when the child starts to walk with the vast majority being of the indirect type (patent

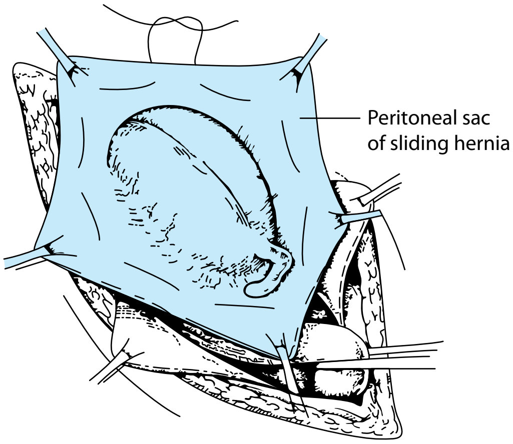


Figure 21.19 A sliding hernia develops when the contents, usually sigmoid colon, descend into the inguinal region. The colon does not lie within the sac but the sac is applied to its surface. The bowel is liable to damage if the sac is fully dissected from the surface and is best managed by a plicating suture after which the bowel can be returned to the abdominal cavity. [Derived from Nyhus LM, London RE (eds). *Hernia*, 2nd edn. Philadelphia, PA: Lippincott, 1978.]

processus vaginalis). Between 10% and 20% of children will develop a hernia on the other side; in about 50%, this peritoneal sac will be present at the time of operation on the presenting side. Attempts to detect these occult hernias have included contrast herniography and laparoscopy. There is no indication for routine exploration of the contralateral side. Direct inguinal hernia is extremely rare in children.

The history of a lump appearing in the groin is usually obtained from the mother, who notices the hernia after a period of straining or coughing by the child. Quite frequently the hernia cannot be demonstrated by the medical examiner and all that can be felt is a silky sensation on palpation of the spermatic cord as the layers of the processus vaginalis move under the examining fingers. Unfortunately, as many of the signs are ignored, there is a tendency for the hernia to become obstructed and strangulated by the time of presentation. This being the case, all children with a strong history should have elective exploration of the affected side. The operation consists of transection and high ligation of the processus vaginalis (herniotomy). Repair of the inguinal canal is not necessary. Differential diagnosis at this age includes undescended testis, hydrocele of the cord and torsion of the testis.

Repair of inguinal hernias

In general, if the patient is fit, an inguinal hernia should be repaired surgically. There are however situations when non-operative management is sensible. This conservative management should only be considered if the hernia is easily reducible and the patient has significant comorbid disease. Small indirect inguinal hernias are controlled by a spring truss, but large indirect and direct hernias require a large pad and firm belt. In all other patients surgical repair should be recommended.

The various repair procedures fall into two categories: fascial repairs (Bassini, Bassini with Tanner's slide, McVay, Ferguson, Shouldice) and tension-free prosthetic mesh repairs, and these may be performed by the anterior open approach or laparoscopically/endoscopically (Box 21.1).

BOX 21.1 Inguinal hernia repair procedures

- Fascial repairs
  - Bassini
  - Bassini with Tanner's slide
  - Shouldice
  - Ferguson
  - McVay
- Tension-free prosthetic mesh repairs
  - Rives–Stoppa
  - Lichtenstein
  - Laparoscopic TEP (Total Extra Peritoneal)
  - Laparoscopic TAPP (Trans-Abdominal PrePeritoneal)

The fascial repairs are the oldest. Their only advantage is the avoidance of prosthetic material, which may become infected, but they carry the highest incidence of recurrence, particularly the Bassini operation since the repair cannot be effected without tension. In practice, infection of the prosthetic mesh, requiring removal, has proved to be a rare occurrence, and this category of hernia repair operation is much more favoured nowadays in view of the uniformly reported good results and low recurrence rates. Open tension-free mesh repairs can be carried out under local anaesthesia as day cases and this reduces the costs considerably.

The laparoscopic approach has a number of advantages that include less postoperative pain, earlier return to full activity and work, and reduced incidence of persistent groin pain at 1 year. The initial higher morbidity (including major vascular, bowel, bladder and nerve injuries) and high recurrence rates reflected inexperience with the technique as surgeons were not familiar with the anatomy of the posterior abdominal wall as visualized from the peritoneal side. The morbidity and recurrence rates of the laparoscopic versus open tension-free repairs are now equivalent. The residual disadvantages are increased hospital (but not total) costs and the need for general or epidural anaesthesia. There is now good evidence that the laparoscopic TEP operation gives the best results in patients with large bilateral and recurrent hernias. The recurrence rate after primary laparoscopic TEP repairs at 3 years is under 1%.

Uniformly excellent results have been reported consistently with the open Lichtenstein repair using polypropylene (Marlex) mesh under local anaesthesia. At 5 years, this procedure in hernia centres has a recurrence rate of 0.1%. Most of the recurrences occur at the pubic tubercle, usually because the mesh used was too small. The tension-free mesh plug repair was introduced in the late 1980s. A preformed mesh plug is used to fill and expand extraperitoneally occluding the defect. In large hernias, the plug is overlaid with an appropriately sized sheet of mesh. The technique is simple and entails minimal dissection, especially in direct hernias. The reported recurrence rates vary from 1% to 3%. In a prospective randomized study comparing open mesh plug repair with laparoscopic TEP, the overall recurrence rate was similar (2.5% for the TEP versus 3% for the mesh–plug hernioplasty). However, patients undergoing the laparoscopic repair required less narcotic analgesic medication and returned to full activity 1 week sooner than the open surgery group. There were no major postoperative complications in either arm in this study but minor morbidity was lower (13%) after laparoscopic TEP than after open mesh plug (23%).

The complications of inguinal hernia repair are:

- urinary retention, especially in males
- wound infection and haematomas
- scrotal swelling
- orchitis and testicular atrophy
- recurrence
- iatrogenic bladder, bowel, nerve and vascular injuries
- chronic groin pain.

Recurrent inguinal hernias

The recurrence rates vary widely, but irrespective of approach, the lowest recurrence rates are encountered after tension-free prosthetic mesh or plug repairs. The majority of recurrence rates occur within the first 2–3 years of the repair. Acceptable recurrence rates should be <3% at 5 years. A Swedish study on a large cohort ($n = 1232$) of patients aged 15–80 years operated upon for inguinal or femoral hernia involving several surgeons/hospitals demonstrated the positive effect of audit (closing the loop) over a 10 year period. The recurrence rate decreased from 18% in 1984–6 to 3% in 1993–4. During this interval, there was a corresponding decline in the reoperation rate for recurrence at 3 years from 10.8% to 2.2%.

The surgical treatment of recurrent inguinal hernias is less effective and the risk of further recurrence is higher than after first time repair. It is now generally agreed that all recurrent inguinal hernias require some form of tension-free prosthetic mesh repair (open or laparoscopic). It is essential that the size of the mesh used is large enough to overlap the defect by a significant margin. The risk of further recurrence depends on the technique and the number of previous repairs. Thus, in a large reported series, patients with a first-time recurrence had recurrence rates of 2% as opposed to 9% in patients who had undergone two or more prior repairs. The morbidity of recurrent hernia repair is higher and includes wound haematoma, scrotal oedema, temporary pain at the wound site, paraesthesiae, injury of the ilioinguinal nerve and femoral hernia, although the overall morbidity can be low with good surgical technique.

Femoral hernias

The pathogenesis of femoral hernias is now thought to be related to the mode of insertion of the fibres of the transversus abdominis and its investing sheath into the superior pubic ramus and develops in two stages. If the insertion of the transversus abdominis fibres on the superior pubic ramus is through a narrow band, a cone-shaped defect overlying the femoral ring (the femoral cone) results. Initially, preperitoneal fat with or without a sac enters this femoral cone as a result of increased abdominal pressure. This is the asymptomatic stage I (internal) femoral hernia that can only be detected if the preperitoneal space is explored during inguinal herniorrhaphy. In time the fatty contents of the femoral cone exit from the narrow distal orifice when a stage II (external) symptomatic hernia results. As the hernia extends downwards, the sac is turned forwards through the cribriform fascia and may then turn upwards to overlie the inguinal ligament, when it may be mistaken for an inguinal hernia. The hernia may, however, remain quite small and be invisible or scarcely palpable in obese patients. Femoral hernia is particularly prone to incarceration and strangulation. By the time of diagnosis 16% of stage II femoral hernias are irreducible and 25–40% present with incarceration/strangulation. Of the emergency group up to 40% will have strangulation of the hernial sac contents (omentum, small bowel, vermiform appendix) requiring excision.

Femoral hernias can occur at any age with peak incidence in the fifth and sixth decades and are significantly commoner in females especially if multiparous (female to male ratio 4:1). The higher incidence on the right side is inexplicable unless the right leg being in use more often than the left in severe exercise is the reason. Femoral hernias are exceedingly rare in children. A small stage II femoral hernia may be difficult to diagnose especially in obese women. The lump may not be easily palpable and the cough impulse difficult to elicit. On other occasions, the nodule, typically below and lateral to the pubic tubercle, may be difficult to differentiate from a lymph node. In such cases, contrast herniography or cross-sectional imaging with CT or MRI can be extremely useful.

Treatment of femoral hernias

All require surgical repair because of risk of obstruction/strangulation. The surgical approach varies with the presentation. In the elective situation, the surgical approach may be from below the inguinal ligament or through the inguinal canal.

Infrainguinal operations

In the classic low approach, the hernial sac is isolated through an incision below the inguinal ligament. The sac is opened and emptied, with care taken to avoid injury to the bladder wall, which may be close to the medial side of the sac. The peritoneum is closed above the neck of the sac and the stump returned to the abdomen. The femoral canal is repaired by interrupted non-absorbable sutures passing from the undersurface of the inguinal ligament to Cooper's ligament behind, or by the insertion of a rolled-up mesh plug. The repair with this cylindrical mesh prosthesis inserted into the femoral canal gives better results than the classic low fascial repair and is favoured nowadays. Alternatively a mesh repair can be effected laparoscopically using the total extraperitoneal approach.

Transinguinal approach

The inguinal canal is opened anteriorly and then the neck of the femoral hernia exposed by incising the posterior inguinal canal wall. The sac is open and the contents are reduced, after which the peritoneum is closed and a tension-free mesh repair is effected of the posterior wall of the inguinal canal, ensuring this is of adequate size and thus overlaps the pubic tubercle.

Obstructed/strangulated femoral hernia

Although both the above can be used in patients with obstructed femoral hernias, the preperitoneal approach of McEvedy is recommended, especially in the presence of strangulation of the contents of the hernial sac because this gives immediate access to the peritoneal cavity. The skin incision may be longitudinal or transverse over the lower abdomen but above the inguinal canal. The musculoaponeurotic layer is divided lateral to the rectus abdominis, and the extraperitoneal space of the lower abdomen is entered. The hernia is usually easily reducible from above but it may not be. In this case the incision is enlarged and the peritoneal cavity entered above the sac.

Other hernias

Richter's or Richter–Littre hernia

In 1700, the French surgeon Alexandre de Littre described a hernia in which only the antimesenteric part of the bowel was inside the hernial sac. Subsequently, in 1777, Richter described 'the intestinal wall hernia', in which the antimesenteric border of the bowel was incarcerated within the hernial sac. For this reason Richter's hernia is sometimes referred as Richter–Littre hernia. Richter's hernia can complicate any small hernia but the most commonly involved is femoral hernia in which a knuckle of the small intestine becomes incarcerated. The patient presents acutely with symptoms and signs of intestinal obstruction. The diagnosis is often delayed as the obstruction may be incomplete or the small hernia may be overlooked particularly in obese patients. There are now well-documented cases of Richter's hernia complicating small unrecognized port site wound hernias after laparoscopic surgery. The diagnosis should be considered in all patients who develop acute intestinal obstruction after laparoscopic surgery and is established by cross-sectional CT imaging.

All patients, irrespective of the nature of Richter's hernia, require urgent intervention after resuscitation; if the partial incarceration of the bowel is necrotic, the affected segment of the small intestine is resected.

Littre's hernia

This is defined as any hernial sac that contains a Meckel's diverticulum and may involve inguinal, umbilical, femoral, ventral and lumbar hernias. Littre's hernia is rare, particularly in children, in whom the umbilical variety is the commonest. It may present with evidence of gastrointestinal bleeding, as an incompletely reducible hernia, or acutely with intestinal obstruction and faecal–hernial fistulas. Obstructed/strangulated Littre's hernia usually presents in adult patients with intestinal obstruction. Preoperative diagnosis is rare with the vast majority being recognized during emergency surgery. The recommended treatment is resection of the Meckel's diverticulum from within the opened hernial sac but extension of the wound is needed when the adjacent loop of small intestine is infarcted, usually as the result of a volvulus.

Obturator (pelvic floor) hernia

Herniation through the obturator foramen is a rare clinical entity occurring most commonly in elderly thin (average body weight 35–40 kg) multiparous women and presenting with acute small bowel obstruction. In one report of a large series one-third of the patients were admitted from homes for elderly people and were either bed-ridden or wheelchair-bound. The vast majority of patients with obstructed obturator hernia are high-risk patients and this together with the delayed diagnosis and intervention accounts for the high morbidity and mortality (15–25%) because of the frequent presence of infarcted bowel (60–75%). The preoperative diagnosis of obturator hernia is difficult since there are no specific clinical features although the Howship–Romberg sign may be positive, but a groin mass is rarely found on physical examination. Recently, pelvic CT has been shown to provide the diagnosis in suspect cases, and may be indicated in elderly patients with mechanical intestinal obstruction of uncertain origin. However, the correct preoperative diagnosis does not appear to influence the outcome, with survival being determined by early surgical intervention.

Repair of the hernial defect is difficult because adjacent tissues are not easily mobilized and thus requires a polypropylene mesh placed in the preperitoneal space and sutured to Cooper's ligament. Recently there have been a few reports on the laparoscopic repair of obturator hernia but the experience is limited. This approach is inadvisable in the presence of infarcted bowel.

Spigelian hernia

This is an interesting hernia that is probably more common than the number of reported cases suggests. The herniation occurs close to the linea semilunaris (Spigelian line). This marks the transition between the muscular and aponeurotic part of the transversus abdominis muscle (Spigelian aponeurosis) at the edge of the rectus abdominis on either side, and extends from the costal margin to the pubic tubercle. The herniation occurs through a defect in the Spigelian aponeurosis between the linea semilunaris and the lateral edge of the rectus (Figure 21.20). As the hernia is covered by the external oblique aponeurosis, it is initially intramural and may not be palpable externally, i.e. the sac lies between the internal and external oblique muscles (85%) but may penetrate through the external oblique layer as it enlarges (15%). The hernial sac may be empty or contain small bowel, omentum and, more rarely, caecum or sigmoid colon. Although very rare, congenital Spigelian hernia has been reported in children and these cases may be associated with an undescended testis or an ipsilateral mediastinal neuroblastoma in which muscle atrophy caused by the neuropathy of the ninth to twelfth intercostal nerves is thought to be responsible for the hernia.

The mean age at diagnosis is 60 years and the hernia can occur on either side with equal frequency. The clinical presentation

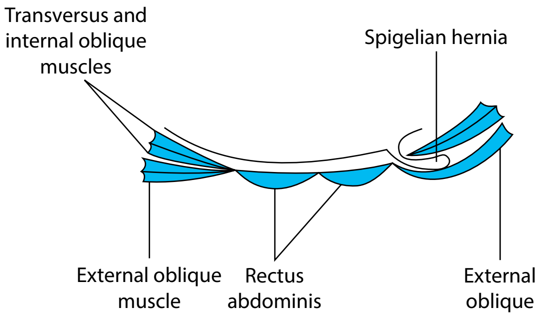


Figure 21.20 Spigelian hernia. The herniation occurs through a defect in the Spigelian aponeurosis between the linea semilunaris and the lateral edge of the rectus.

varies depending on the contents of the hernia sac and the size/type of herniation. Pain is the most common symptom but varies in intensity and nature. On physical examination, the commonest findings are a palpable hernia and a palpable hernial defect. Some patients (20–25%) present acutely with a tender irreducible mass or intestinal obstruction caused by incarceration of a loop of intestine.

Although large easily palpable Spigelian hernias do not pose diagnostic problems, small hernias are often overlooked. Persistent point tenderness along the Spigelian aponeurosis with associated spasm of the abdominal wall should suggest the diagnosis in these cases. Ultrasound scanning or CT will identify these small impalpable hernias.

The treatment of Spigelian hernia is surgical. A grid iron incision is favoured by many surgeons in patients with palpable hernias. In patients with non-palpable hernias the preperitoneal dissection is carried out through a vertical incision, which permits good exposure. The vertical approach is also recommended in patients requiring emergency surgery because of intestinal obstruction as it enables an exploratory laparotomy. Repair can be either fascial or by prosthetic mesh sheet or plug, although as with other hernias tension-free prosthetic repair is favoured nowadays. The repair of uncomplicated Spigelian hernias can be performed laparoscopically. Irrespective of approach, the reported recurrence rates after surgical repair are low.

GUIDE TO FURTHER READING

Adam U, Mack D, Forstner R, *et al.* Conservative treatment of acute Ormond's disease. *Tech Urol* 1999;**5**:54–56.

Athanassiadi K, Kalavrouziotis G, Athanassiou M, *et al.* Blunt diaphragmatic rupture. *Eur J Cardiothorac Surg* 1999;**15**:469–474.

Ballo MT, Zagars GK, Pollack A, *et al.* Desmoid tumor: prognostic factors and outcome after surgery, radiation therapy, or combined surgery and radiation therapy. *J Clin Oncol* 1999;**17**:158–167.

Barbalias GA, Liatsikos EN. Idiopathic retroperitoneal fibrosis revisited. *Int Urol Nephrol* 1999;**31**:423–429.

Bauer JJ, Harris MT, Gorfine SR, Kreel I. Rives-Stoppa procedure for repair of large incisional hernias: experience with 57 patients. *Hernia* 2002;**6**:120–123.

Cheng IK, Fang GX, Chau PY, *et al.* A randomized prospective comparison of oral levofloxacin plus intraperitoneal (IP) vancomycin and IP netromycin plus IP vancomycin as primary treatment of peritonitis complicating CAPD. *Perit Dial Int* 1998;**18**:371–375.

Church J, Berk T, Boman BM, *et al.* Collaborative Group of the Americas on Inherited Colorectal Cancer. Staging intraabdominal desmoid tumors in familial adenomatous polyposis: a search for a uniform approach to a troubling disease. *Dis Colon Rectum* 2005;**48**:1528–1534.

Clark SK, Smith TG, Katz DE, *et al.* Identification and progression of a desmoid precursor lesion in patients with familial adenomatous polyposis. *Br J Surg* 1998;**85**:970–973.

Clark SK, Neale KF, Landgrebe JC, Phillips RK. Desmoid tumours complicating familial adenomatous polyposis. *Br J Surg* 1999;**86** 1185–1189.

de Perrot M, Rostan O, Morel P, Le Coultre C. Abdominal lymphangioma in adults and children. *Br J Surg* 1998;**85**:395–397.

Ellis H, Moran BJ, Thompson JN, *et al.* Adhesion-related hospital readmissions after abdominal and pelvic surgery: a retrospective cohort study. *Lancet* 1999;**353**:1476–1480.

Fanburg-Smith JC, Miettinen M. Liposarcoma with meningothelial-like whorls: a study of 17 cases of a distinctive histological pattern associated with dedifferentiated liposarcoma. *Histopathology* 1998;**33**:414–424.

Guarner C, Soriano G. Spontaneous bacterial peritonitis. *Semin Liver Dis* 1997;**17**:203–217.

Guarner C, Sola R, Soriano G, *et al.* Risk of a first community-acquired spontaneous bacterial peritonitis in cirrhotics with low ascitic fluid protein levels. *Gastroenterology* 1999;**117**:414–419.

Hauser H, Mischinger HJ, Beham A, *et al.* Cystic retroperitoneal lymphangiomas in adults. *Eur J Surg Oncol* 1997;**23**:322–326.

Heling KS, Chaoui R, Hartung J, *et al.* Prenatal diagnosis of congenital neuroblastoma. Analysis of 4 cases and review of the literature. *Fetal Diagn Ther* 1999;**14**:47–52.

Hemsley C, Eykyn SJ. Pneumococcal peritonitis in previously healthy adults: case report and review. *Clin Infect Dis* 1998;**27**:376–379.

Herman K, Kusy T. Retroperitoneal sarcoma: the continued challenge for surgery and oncology. *Surg Oncol* 1998;**7**:77–81.

Hinson FL, Ambrose NS. Pseudomyxoma peritonei. *Br J Surg* 1998;**85**:1332.

Huddy CL, Boyd PA, Wilkinson AR, Chamberlain P. Congenital diaphragmatic hernia: prenatal diagnosis, outcome and continuing morbidity in survivors. *Br J Obstet Gynaecol* 1999;**106**:1192–1196.

Janssen-Heijnen ML, Damhuis RA, Klinkhamer PJ, *et al.* Increased but low incidence and poor survival of malignant mesothelioma in the southeastern part of The Netherlands since 1970: a population-based study. *Eur J Cancer* 1999;**1/8**:311–314.

Jaquet P, Sugarbaker PH. Clinical research methodologies in diagnosis and staging of patients with peritoneal carcinomatosis. In: Sugarbaker PH (ed.) *Peritoneal Carcinomatosis: Principles of Management*. Boston, MA: Kluwer Academic, 1996:359–374.

Khan MA, Lo AY, Vande Maele DM. Paraduodenal hernia. *Am Surg* 1998;**64**:1218–1222.

Khoury N. A randomized prospective controlled trial of laparoscopic extraperitoneal hernia repair and mesh-plug hernioplasty: a study of 315 cases .*J Laparoendosc Adv Surg Tech A* 1998;**8**:367–372.

Killeen KL, Mirvis SE, Shanmuganathan K. Helical CT of diaphragmatic rupture caused by blunt trauma. *Am J Roentgenol* 1999;**173**:1611–1616.

Lackner H, Urban C, Kerbl R, *et al.* Noncytotoxic drug therapy in children with unresectable desmoid tumors. *Cancer* 1997;**80**:334–340.

Leber GE, Grab JL, Alexander AI, *et al.* Long-term complications associated with prosthetic repair of incisional hernias. *Arch Surg* 1998;**133**:378–382.

Levy AD, Cantisani V, Miettinen M. Abdominal lymphangiomas: imaging features with pathologic correlation. *AJR Am J Roentgenol* 2004;**182**:1485–1491.

Lewis JJ, Boland PJ, Leung DH, *et al.* The enigma of desmoid tumors. *Ann Surg* 1999;**229**:866–872.

Li C, Bapat B, Alman BA. Adenomatous polyposis coli gene mutation alters proliferation through its beta-catenin-regulatory function in aggressive fibromatosis (desmoid tumor). *Am J Pathol* 1998;**153**:709–714.

Mann GB, Lewis JJ, Brennan MF. Adult soft tissue sarcoma. *Aust NZ J Surg* 1999;**69**:336–343.

Merchant NB, Lewis JJ, Woodruff JM, *et al.* Extremity and trunk desmoid tumors: a multifactorial analysis of outcome. *Cancer* 1999;**86**:2045–2052.

Merrot T, Chaumoitre K, Simeoni-Alias J, *et al.* Abdominal cystic lymphangiomas in children. Clinical, diagnostic and therapeutic aspects: apropos of 21 cases. *Ann Chir* 1999;**53**:494–499.

Mongero LB, Beck JR, Kroslowitz RM, *et al.* Treatment of primary peritoneal mesothelioma by hyperthemic intraperitoneal chemotherapy. *Perfusion* 1999;**14**:141–145.

Nathens AB, Rotstein OD, Marshall JC. Tertiary peritonitis: clinical features of a complex nosocomial infection. *World J Surg* 1998;**22**:158–163.

Ozenc A, Ozdemir A, Coskun T. Internal hernia in adults. *Int Surg* 1998;**83**:167–170.

Ozener C, Kiris S, Lawrence R, *et al.* Potential beneficial effect of tamoxifen in retroperitoneal fibrosis. *Nephrol Dial Transplant* 1997;**12**:2166–2168.

Park BJ, Alexander HR, Libutti SK, *et al.* Treatment of primary peritoneal mesothelioma by continuous hyperthermic peritoneal perfusion (CHPP). *Ann Surg Oncol* 1999;**6**:582–590.

Rai S, Chandra SS, Smile SR. A study of the risk of strangulation and obstruction in groin hernias. *Aust NZ J Surg* 1998;**68**:650–654.

Raj J, Urban LM, ReMine SG, Raj PK. Laparoscopic management of pseudomyxoma peritonei secondary to adenocarcinoma of the appendix. *J Laparoendosc Adv Surg Tech A* 1999;**9**:299–303.

Ramachandran CS. Umbilical hernial defects encountered before and after abdominal laparoscopic procedures. *Int Surg* 1998;**83**: 171–173.

Ramirez OM, Ruas E, Dellon AL. Components separation method for closure of abdominal-wall defects: an anatomic and clinical study *Plast Reconstr Surg* 1990;**86**:519–526.

Robbins AW, Rutkow IM. Mesh plug repair and groin hernia surgery. *Surg Clin North Am* 1998;**78**:1007–1023.

Sanchez-Bustos F, Ramia JM, Fernandez Ferrero F. Prosthetic repair of femoral hernia: audit of long term follow-up. *Eur J Surg* 1998;**164**:191–193.

Sandblom G, Gruber G, Kald A, Nilsson E. Audit and recurrence rates after hernia surgery. *Eur J Surg* 2000;**166**:154–158.

Schumpelick V, Steinau G, Schluper I, Prescher A. Surgical embryology and anatomy of the diaphragm with surgical applications. Surg *Clin North Am* 2000;**80**:213–239.

Shestak KC, Edington HJ, Johnson RR. The separation of anatomic components technique for the reconstruction of massive midline abdominal wall defects: anatomy, surgical technique, applications, and limitations revisited. *Plast Reconstr Surg* 2000;**105**:731–738.

Shin HJ, Sneige N. Epithelial cells and other cytologic features of pseudomyxoma peritonei in patients with ovarian and/or appendiceal mucinous neoplasms: a study of 12 patients including 5 men. *Cancer* 2000;**90**:17–23.

Skandalakis LJ, Androulakis J, Colborn GL, Skandalakis JE. Obturator hernia. Embryology, anatomy, and surgical applications. *Surg Clin North Am* 2000;**80**:71–84.

Skapek SX, Hawk BJ, Hoffer FA, *et al.* Combination chemotherapy using vinblastine and methotrexate for the treatment of progressive desmoid tumor in children. *J Clin Oncol* 1998;**16**:3021–3027.

Sugarbaker PH. Management of peritoneal-surface malignancy: the surgeon's role. *Langenbecks Arch Surg* 1999;**384**:576–587.

Tejpar S, Nollet F, Li C, *et al.* Predominance of beta-catenin mutations and beta-catenin dysregulation in sporadic aggressive fibromatosis (desmoid tumor). *Oncogene* 1999; **18**:6615–6620.

Van der Kabben AA, Dijkstra M, Nieuwenhuizen MM, *et al.* Morbidity and mortality of inadvertent enterotomy during adhesiotomy. *Br J Surg* 2000;**87**:467–471.

von Fritschen U, Malzfeld E, Clasen A, Kortmann H. Inflammatory abdominal aortic aneurysm: a postoperative course of retroperitoneal fibrosis. *J Vasc Surg* 1999;**30**:1090–1098.

Vrijland W, Jeekel J, Steyerberg EW, *et al.* Intraperitoneal polypropylene mesh repair of incisional hernia is not associated with enterocutaneous fistula. *Br J Surg* 2000;**87**:348–352.

White TJ, Santos MC, Thompson JS. Factors affecting wound complications in repair of ventral hernias. *Am Surg* 1998;**64**:276–280.

Wirtzfeld DA, Rodriguez-Bigas M, Weber T, Petrelli NJ. Disseminated peritoneal adenomucinosis: a critical review. *Ann Surg Oncol* 1999;**6**:797–801.

CHAPTER 22

Disorders of the oesophagus

GEORGE B. HANNA, SAMI M. SHIMI AND SIR ALFRED CUSCHIERI

Anatomy and physiology of the oesophagus

Embryology of the oesophagus

The oesophagus starts to develop in the fourth week of embryonic development from the foregut immediately caudal to the primordial pharynx and extends through the fusiform dilatation in the foregut, this time to become the stomach. The oesophagus is short initially, but it elongates rapidly owing to the growth and descent of the heart and lungs, reaching its final relative length by the seventh week. As the laryngotracheal groove grows from the ventral wall of the primordial pharynx, longitudinal tracheo-oesophageal folds grow and approach each other to fuse and form a septum between the developing trachea and oesophagus. Incomplete fusion of the tracheo-oesophageal folds results in a defective tracheo-oesophageal septum and fistula formation. By the fifth week of development the oesophageal epithelium is two cells thick and is composed of columnar cells, which develop cilia by 10 weeks. The epithelium lining proliferates and partly or completely obliterates the lumen by the eighth week of development and large vacuoles appear in the centre. In succeeding weeks, the vacuoles coalesce and the oesophageal lumen recanalizes but with a multilayered ciliated epithelium. During the fourth month this epithelium finally becomes replaced with the stratified squamous epithelium that characterizes the mature oesophagus. Failure of recanalization of the oesophagus in this period results in oesophageal atresia and possibly oesophageal stenosis. Surrounding the oesophageal epithelium, layers of muscle begin to differentiate from the mesoderm. The inner circle of the muscular layer is recognizable by 5 weeks of gestation and the outer longitudinal layer of muscle begins to take shape by 8 weeks of gestation. The striated muscle in the superior third of the oesophagus is derived from mesenchyme in the caudal pharyngeal arches, whereas the smooth muscle in the inferior third of the oesophagus develops from the surrounding splanchnic mesenchyme. Both types of muscle are innervated by branches of the vagus nerve, which supply the caudal pharyngeal arches. A dorsal mesentery more developed inferiorly connects the oesophagus to the developing aorta and may, on occasion, have smooth muscle remnants in adult life. Vagal trunks run alongside the oesophagus, but as the stomach rotates to the right, the right vagus assumes a posterior position and the left trunk crosses the lower oesophagus to lie anterior to the oesophagogastric junction. The oesophagus attains its final length at the seventh week of gestation, having a length at birth of 8–10 cm.

Anatomy of the oesophagus

The oesophagus is a hollow muscular tube, which is about 25 cm long, and connects the pharynx to the stomach. It commences in the neck, level with the lower border of the cricoid cartilage (sixth cervical vertebra) and descends mainly anterior to the vertebral column traversing the diaphragm at the level of the tenth thoracic vertebra, ending in the abdomen at the cardiac orifice of the stomach, at the level of the eleventh thoracic vertebra. Topographically the oesophagus is generally vertical but has two shallow curves. Immediately below the pharynx it is a midline structure but inclines to the left as far as the root of the neck, gradually returning to the midline near the fifth thoracic vertebra and deviates to the left again at the seventh thoracic vertebra finally turning anteriorly as it traverses the diaphragm. The oesophagus also curves in a coronal plane, to follow the cervical thoracic curvatures of the vertebral column. The surgical relevance of these deviations is that the cervical oesophagus is best approached from the left side of the neck, and the thoracic portion through the right side of the thorax except the lower third (below the thoracic arch), which is more accessible from the left thorax.

The oesophagus is arbitrarily divided into three segments:

- *The cervical part* (5 cm long) is behind the trachea and attached to it by loose areolar tissue. It lies in front of the prevertebral fascia separated from it by loose areolar tissue. The recurrent laryngeal nerves ascend on each side in the groove between the trachea and the oesophagus. The left recurrent laryngeal nerve is positioned closer to the oesophagus than the right one as it ascends along the oesophagus for a larger distance. The cervical oesophagus commences with cricopharyngeus muscle at the inferior portion of the inferior pharyngeal constrictor, clearly identified by its transverse fibres. Just above cricopharyngeus, the transition between the oblique fibres of the inferior constrictor muscle and the transverse fibres of cricopharyngeus creates a point of potential weakness termed Killian's dehiscence, which is the site of origin of a pharyngo-oesophageal (Zenker's) diverticulum (Figure 22.1). Cricopharyngeus muscle fibres blend into the longitudinal and circular muscles of the cervical oesophagus.
- *The thoracic oesophagus* runs in the superior mediastinum between the trachea and the vertebral column, and passes behind and to the right of the aortic arch to descend in the posterior mediastinum along the right side of the descending thoracic aorta as it proceeds behind the pericardium overlying the left atrium. The oesophagus then deviates further to the left and anteriorly, entering the oesophageal diaphragmatic hiatus at the level of the tenth thoracic vertebra. On either side the thoracic oesophagus is bound by the parietal pleura. Clinically, the thoracic oesophagus is divided into three parts. The upper thoracic oesophagus extends from the cricopharyngeus to the level of the carina. The middle thoracic oesophagus extends from the level of the carina to halfway between the carina and the oesophagogastric junction and the lower thoracic oesophagus from halfway between the carina and the oesophagogastric junction to include the lower third of the oesophagus. Oncologically the thoracic oesophagus is divided into the supracarina oesophagus (upper oesophagus) and the infracarina oesophagus (middle and lower oesophagus).
- *The abdominal oesophagus* emerges from the right diaphragmatic crus slightly to the left of the midline at the level of the tenth thoracic vertebra. It forms an inverted cone (1–2 cm in length) which curves sharply to the left and its base continues with the gastric cardiac orifice. It is covered by peritoneum on its front and left side, and the peritoneum reflected from its posterior surface to the diaphragm is part of the gastrophrenic ligament. Behind it is the left crus of the diaphragm.

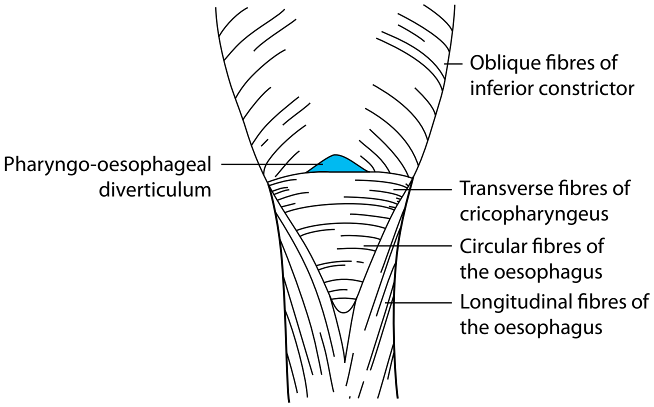

Oblique fibres of inferior constrictor
Pharyngo-oesophageal diverticulum
Transverse fibres of cricopharyngeus
Circular fibres of the oesophagus
Longitudinal fibres of the oesophagus

Figure 22.1 Diagrammatic representation of the posterior aspect of the upper oesophagus and pharynx showing Killian's dehiscence as the longitudinal fibres of the oesophagus sweep anteriorly. The pharyngo-oesophageal diverticulum emerges between the oblique fibres of the inferior constrictor and the transverse fibres of the cricopharyngeus.

Oesophageal attachments

The oesophagus is loosely bound to adjacent structures by fibroareolar tissue throughout its course except at the upper and lower ends. Superiorly, the longitudinal muscle fibres of the oesophagus are inserted into the cricoid cartilage. The lower attachments consist of serous reflections and the phreno-oesophageal membrane. The subdiaphragmatic pleural reflection is continuous with the mediastinal pleura and is separated from the lower segment of the oesophagus by a condensation of the endothoracic fascia which constitutes the phreno-oesophageal membrane. This important fibroelastic membrane fixes the lower gullet but permits its continuous vertical displacement as occurs with respiration. The phreno-oesophageal membrane consists of a superior and an inferior limb. The latter is inserted into the cardia and the superior limb into the lower 3 cm of the thoracic oesophagus. The fibres of the membrane are disposed in bundles and lamelli and are inserted deeply into the oesophageal walls, some reaching the submucous layer. Approximately 40–60% of the fibres are made of elastin. The membrane has both strength and resilience, which are necessary to cope with the continuous movement of the hiatus during life. The phreno-oesophageal membrane is easily identified during mobilization of the oesophagus during laparoscopic antireflux surgery.

Structure of the oesophagus

The oesophagus consists of four layers. The fibrous adventitia is irregular, and consists of loose, areolar connective tissue containing elastin fibres. The looseness permits considerable movement of the oesophagus during swallowing. The fibres penetrate and surround the fasciculae of muscle in the deeper layers. The muscular layer is composed of an outer thicker longitudinal and inner circular layer. The longitudinal fibres surround the whole length of the oesophagus with a continuous coat except posterosuperiorly where the longitudinal fibres separate and sweep round to the anterior aspect of the oesophagus before their insertion into the posterior aspect of the cricoid cartilage. The V-shaped interval between these fasciculae is filled by cricopharyngeus above and circular muscle fibres below. The potential weak area posteriorly may become the site of acquired pulsion diverticulae (Zenker's above and Laimer's below). A natural constriction occurs at this point due to the presence of the hypopharyngeal fold and the cricopharyngeus muscle. Accessory slips of non-striated muscle sometimes pass between the oesophagus and left pleura or the root of the left principal bronchus, trachea, pericardium or aorta. The circular fibres are continuous with those of cricopharyngeus superiorly and with the oblique gastric muscle fibres inferiorly. The two muscle layers of the upper quarter of the oesophagus are striated. Below this there is a gradual replacement with smooth muscle. Although similar in appearance to striated skeletal muscle it is not under voluntary control but rather under autonomic nervous control. At the lower end of the oesophagus, the circular muscle layer is thickened but a definite anatomical sphincter is not present. The external longitudinal muscle of the oesophagus continues longitudinally along the gastric greater and lesser curvatures. These muscle bundles turn upward toward the fundus, interlacing with fibres of the internal muscle layer. The circular muscle layer of the lowermost portion of the oesophagus becomes semicircular (clasps) and continues down the lesser curvature aspect of the cardia to insert in the submucosal connective tissue on the opposite side. Gastric sling fibres are also semicircular in the opposite direction and at an oblique angle to the posterior attachments of the oesophageal semicircular muscles.

The submucosa is very loose in order to permit dilatation of the oesophagus during swallowing. It loosely connects mucous and muscular layers and contains large blood vessels, nerves and mucous glands. The mucosa is made of non-keratinized squamous epithelium, which is arranged in longitudinal folds especially at the lower end where the oesophageal mucosal folds form a rosette. The mucosal layer consists of the lining epithelium, connective tissue with papilli (lamina propria) and non-striated muscularis mucosa. At the upper oesophagus the muscularis mucosa is absent or sparse and becomes a considerable stratum below this. At intermediate levels its fascicles are mainly longitudinal but become more plexiform towards the gastro-oesophageal junction (GOJ). There are small oesophageal glands of compound racemose mucous type in the submucosa deep to the muscularis mucosa each with a long duct traversing it and the other layers of the mucosa. Glands of the abdominal portion of the oesophagus resemble gastric cardia glands and lie superficial to the muscularis mucosa.

The cardio-oesophageal junction

This consists of the supradiaphragmatic portion, the inferior oesophageal constriction, the vestibule, and the cardia (Figure 22.2). Radiologically the supradiaphragmatic portion consists of an ampulla and the empty segment just distal to it. The ampulla is not an anatomical dilatation and is caused by the primary peristaltic wave (during a contrast examination) acting in conjunction with the negative intrathoracic pressure, which momentarily expands the segment just before the lower oesophageal sphincter (LDS) relaxes. The inferior oesophageal constriction consists of a concentric narrowing of the oesophageal lumen at the level of the diaphragmatic hiatus. On average, it is situated some 2 cm from the cardio-oesophageal junction. It is not synonymous with the LOS although the sphincter region extends to include this area. The longitudinal oesophageal mucosal folds are very prominent inside the inferior constriction but disappear readily when the oesophagus is distended. The intra-abdominal segment of the oesophagus is also known as the vestibule. It is often described as an inverted funnel or cone which inclines to the left before joining the stomach at an angle (cardiac angle or angle of His). When viewed endoscopically from within the stomach it forms a well-marked ridge at the left margin of the gastro-oesophageal junction, which is occasionally referred to as the incisura. The cardia denotes the junction between the oesophagus and stomach. The only reliable and constant anatomical landmark of this is made by the sling or oblique muscle fibres of the stomach but these cannot be identified

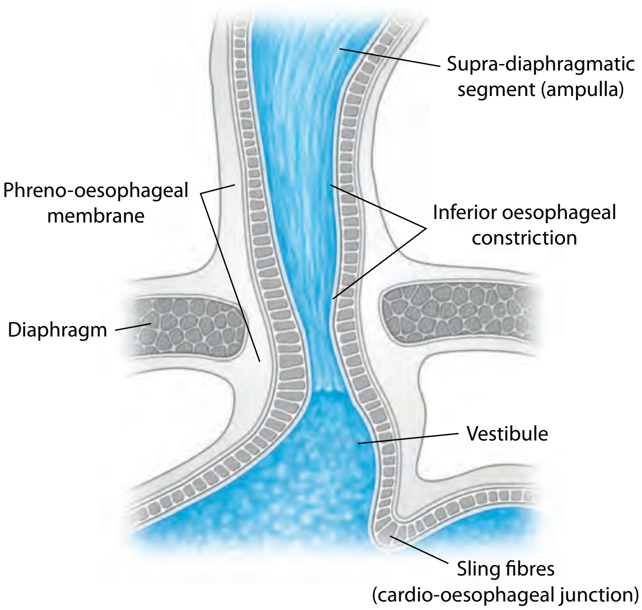

Supra-diaphragmatic segment (ampulla)
Phreno-oesophageal membrane
Inferior oesophageal constriction
Diaphragm
Vestibule
Sling fibres (cardio-oesophageal junction)

Figure 22.2 Diagrammatic representation of the anatomy of the lower oesophagus. (From Hennessey and Cuschieri, *Surgery of the Oesophagus*. Oxford, UK: Butterworth Heineman, 1992, with permission.)

endoscopically. Clinically the term cardia is used to describe the junction between the oesophagus and stomach. It contains the squamocolumnar junction, which forms the serrated z-line marking an abrupt change from the tough smooth pale squamous epithelium of oesophagus to the epithelium of the stomach. A zone of junctional epithelium is interposed between the squamous lining of the oesophagus and the gastric mucosa. It is lined by columnar cells which contain simple tubular mucosal glands which are superficial to the muscularis mucosa and hence resistant to acid and peptic digestion.

Vascular supply, venous and lymphatic drainage

The oesophagus receives a segmental arterial supply from several small arterioles throughout its course. These come from the inferior thyroid, common carotid, costocervical and vertebral arteries in the neck, bronchial and direct aortic branches in the thorax, and branches of the left gastric and left inferior phrenic artery at the lower end including the abdominal portion. The arterioles terminate in a fine capillary network before penetrating the oesophageal muscle layer. After penetrating and supplying the oesophageal muscle layer, the arterioles join a capillary plexus in the submucosa. These capillary networks run longitudinally in the submucosa allowing mobilization of segments of the oesophagus without compromise to the blood supply. The venous drainage is to the inferior thyroid and hypopharyngeal veins in the neck, and the azygos, hemiazygos and intercostal veins in the chest except at the lower end which drains into the left gastric vein (Figure 22.3). The lower oesophagus is clinically the most important site of communication between the portal and systemic venous radicles, being the site of occurrence of varices in the majority of patients with portal hypertension. In the lower oesophagus the veins are mainly situated within the lamina propria whereas in the more proximal oesophagus and stomach they reside in the submucous plane. This is held responsible for the propensity of the lower oesophagus to bleed in variceal portal hypertension. The lymphatics form extensive mucosal (lamina propria), submucosal, muscularis and adventitia plexuses which communicate freely and lymph flows long distances in the large submucosal plexus before passing though the muscular coat to reach the adventitial plexus and draining lymph nodes. These are grouped into three main tiers: the first composed of nodes alongside the oesophagus (paraoesophageal), the second or intermediate group is made up of mediastinal lymph nodes, and the third are the deep cervical, supraclavicular, tracheobronchial and coeliac nodes from above downwards. In general, lymph drainage from the upper two-thirds of the oesophagus proceeds in a proximal direction towards the cervical region, whereas the lower third drains distally to the subdiaphragmatic region and coeliac lymph nodes. In view of the anatomic distribution and communication of oesophageal lymphatics, oncological resection should include 10 cm above and below the tumour. Lymphatic drainage of the GOJ mainly follows the arteries supplying the junction.

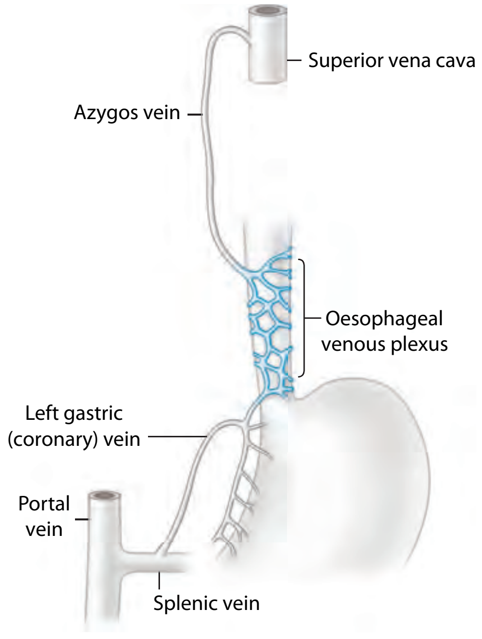


Figure 22.3 Venous drainage of the lower oesophagus.

The thoracic duct arises from the cysterna chyli, which lies in the abdomen to the right of the aorta, approximately at the level of the second lumbar vertebra. The duct enters the chest through the aortic hiatus coursing in the posterior mediastinum to the right of the midline between the aorta and the azygos vein behind the oesophagus. At the level of the fifth thoracic vertebra, it crosses behind the oesophagus to the left under the aortic arch and continues along the left side of the oesophagus ascending behind the left subclavian artery to the base of the neck. There, it curves to the right and caudally to drain into the internal jugular vein near its junction with the left subclavian vein. The majority of people have a single thoracic duct and the remainder have two or more ducts.

Nerve supply of the oesophagus

Two plexuses of nerves in the oesophageal wall (Meissner's plexus in the submucosa and Auerbach's plexus in the muscularis) form networks of multipolar ganglion cells, the processes of which are in contact with one another and receive axons from the vagus. Postganglionic fibres of these plexuses innervate the smooth muscle cells. Postganglionic sympathetic fibres from the prevertebral ganglia enter the plexuses without synapsing to reach the muscles of the blood vessels of the oesophageal wall.

Swallowing and oesophageal motility are co-ordinated by the swallowing centre in the brainstem. This is an area in the medulla oblongata and lower portion of the pons, which is closely associated with the tractus solitarius. It receives and co-ordinates sensory inputs from peripheral mechanoreceptors in the pharynx and oesophagus. Motor impulses that initiate the swallowing reflex are transmitted from the swallowing centre to the pharynx and upper oesophagus via the trigeminal, glossopharyngeal, vagus and hypoglossal nerves.

The nerve supply of the oesophagus is parasympathetic (vagal) and sympathetic. The parasympathetic fibres, which are predominantly motor, are derived from neurones of the vagal motor nuclei and travel in the glossopharyngeal, vagus and recurrent laryngeal nerves terminating in the myenteric plexus. The sensory component is formed by axons of the nodose ganglion, the cells of which are unipolar with their peripheral axons transmitted via the sensory portion of the trigeminal and glossopharyngeal nerves, which terminate in receptors situated in the pharynx and oesophagus. The central axons of the cells of the nodose ganglion communicate with the swallowing centre in the brainstem.

The sympathetic nerve supply consists of preganglionic fibres derived from neurones situated in the intermediolateral columns of spinal cord segments T5 and T6 and terminate in the cervical, thoracic and coeliac ganglia. The postganglionic sympathetic fibres then reach the oesophagus as a periarterial plexus.

The rami reaching the oesophagus communicate with a plexus containing groups of ganglion cells between the two layers of the muscular coat, the myenteric plexus. The submucous plexus of the oesophagus is rather sparse and consists mainly of nerve fibres.

Endoscopic appearance of the oesophagus

The oesophagus proper starts just distal to the cricopharyngeus muscle at around 15 cm from the incisor teeth. The mucosa is pale and lacks lustre. The lumen is collapsed until insufflated with air and then the wall is smooth. There are a number of naturally occurring anatomic narrowings in the oesophagus. The cervical constriction occurs at the level of the cricopharyngeus muscle at approximately 15 cm from the incisor teeth. The next constriction is expected where the oesophagus is crossed by the aortic arch at 22 cm, by the left main bronchus at 27 cm, and where it passes through the diaphragmatic hiatus at the cardio-oesophageal junction at 37–40 cm from the incisors depending on the sex and build of the patient. Although the left atrium is in front of the lower part of the oesophagus below the left main bronchus, it is only when the atrium is enlarged that it indents the oesophagus. At the cardio-oesophageal junction the stratified squamous epithelium is abruptly succeeded by gastric columnar epithelium, the junction is visible as a serrated line with the greyish pink smooth oesophageal mucosa contrasting with the red mamillated gastric mucosa.

Physiology of the oesophagus

The basic function of the oesophagus is to transport swallowed material from the pharynx into the stomach. It has little secretory function and no absorptive function. The pharynx is central to the swallowing complex. Besides other functions the pharynx facilitates the propulsion of food for a few seconds at a time. Conventionally, swallowing has been divided into three stages. The *voluntary stage* initiates the swallowing process. A food bolus is rolled (squeezed) posteriorly into the pharynx by pressure of the tongue upwards and backwards against the palate. Subsequently, the process of swallowing becomes automatic and involuntary. The *pharyngeal stage* constitutes the passage of food through the pharynx into the oesophagus. As food enters the pharynx it stimulates mechanoreceptors around the opening of the pharynx and on the tonsilar pillars. This stimulates a series of automatic pharyngeal muscular contractions. The soft palate is pulled upward to close the posterior layers, preventing reflux of food into the nasal cavities. The palatopharyngeal folds on either side approximate each other to form a sagittal slit allowing properly masticated food to pass with ease and yet impeding the passage of large objects. The vocal cords are strongly approximated and the larynx is pulled upward and posteriorly against the epiglottis, which swings backward over the opening of the larynx. This prevents passage of food into the trachea. The upward movement of the larynx enlarges the opening of the oesophagus and stimulates relaxation of the tonically active upper oesophageal sphincter (UOS) (cricopharyngeus). As the larynx is raised and the UOS is relaxed the entire muscular wall of the pharynx contracts beginning in the superior constrictor and spreading downwards as a rapid peristaltic wave over the middle and interior constrictors and thence into the oesophagus. This peristaltic wave propels the food into the oesophagus. This stage of swallowing is principally a reflex act. It is almost always initiated by voluntary movement of food into the back of the mouth, which in turn elicits the swallowing reflex. The *oesophageal stage* is also involuntary and promotes passage of food from the pharynx to the stomach. Food is propelled down the oesophagus by peristaltic waves. Primary peristalsis is simply a continuation of the peristaltic wave that commenced in the pharynx and spread into the oesophagus during the pharyngeal phase of swallowing. This wave passes all the way from the pharynx to the stomach in approximately 8 seconds. However, food swallowed in the upright position is usually transmitted down the oesophagus in about 5–8 seconds because of the additional effect of gravity. If the primary peristaltic wave fails to move all the food that has entered the oesophagus into the stomach secondary peristaltic waves result from the distension of the oesophagus by the retained food and these continue until all the food has emptied into the stomach. The secondary peristaltic waves are initiated partly by intrinsic neuronal circuits in the myenteric plexus and partly by reflexes that are transmitted through vagal fibres from the brainstem. It is important to note that a denervated oesophagus will continue to generate these secondary peristaltic waves in response to distension. The other types of contraction exhibited by the oesophagus are tertiary contractions which occur spontaneously and are non-propulsive. They are usually encountered in various oesophageal motility disorders but are also observed in apparently healthy individuals especially in older age. These tertiary contractions may be generalized involving the whole of the oesophageal body or localized. They are usually seen on a double contrast barium meal examination.

In between swallows the body of the oesophagus is flaccid and the lumen largely collapsed. However, both the UOS and LOS are tonically contracted. The UOS (cricopharyngeus) is clipped closed by tonic contraction induced through constant discharge from neurones originating in the cranial nuclei. The sphincter remains in tonic contraction with a resting pressure of about 100 mmHg. The physiological function of the sphincter is to prevent passage of air from the pharynx into the oesophagus and reflux of oesophageal contents into the pharynx. This reflex action is essentially protective and is aimed at preventing aspiration. In the resting state the oesophageal body has no motor activity. When stimulated by passage of food or saliva a contraction is initiated in the upper oesophagus, which progresses distally towards the stomach. Oesophageal peristaltic waves travel at 3–4 cm/s, last between 3 and 4.5 seconds, and reach a peak amplitude of 30 mmHg in the upper oesophagus which progressively gets larger as it travels down the oesophagus to around 120 mmHg in the lower oesophagus. In the resting state the pressure within the body of the oesophagus is similar to the intrathoracic pressure, being negative during inspiration and reaches 5 mmHg during expiration. At the lower 2–4 cm of the oesophagus, a high-pressure zone (HPZ) is encountered. This is due to tonic contraction of the LOS (intrinsic component) and to mechanical factors (extrinsic component).

The intrinsic part of the HPZ constitutes the 2–4 cm tonically contracted segment just proximal to the gastro-oesophageal junction. Its normal resting tone ranges from 15 to 25 mmHg relative to intragastric pressure. Three-dimensional mapping of this region shows asymmetry of the pressure profile (Figure 22.4). The segment has a spontaneous tendency to relax periodically at times unrelated to swallowing. These periodic relaxations have been termed 'transient lower oesophageal sphincter relaxations' to distinguish them from relaxations triggered by swallows. They account for the small amount of physiological reflux found in normal subjects. To allow a food bolus to enter the stomach, the LOS relaxes for 5–10 seconds and then tonic contraction returns. Relaxation is probably due to non-adrenergic non-cholinergic neurotransmitters such as vasoactive intestinal polypeptide (VIP) and nitric oxide (NO). The resting tone is largely dependent on the intrinsic muscular activity because it persists after neural input is abolished by treatment with neurotoxins. During fasting the LOS has cyclic phasic contractile activity synchronous with phases 2 and 3 of the interdigestive motor complex. This is probably regulated by motilin, which acts on the LOS, by preganglionic stimulation of cholinergic nerves.

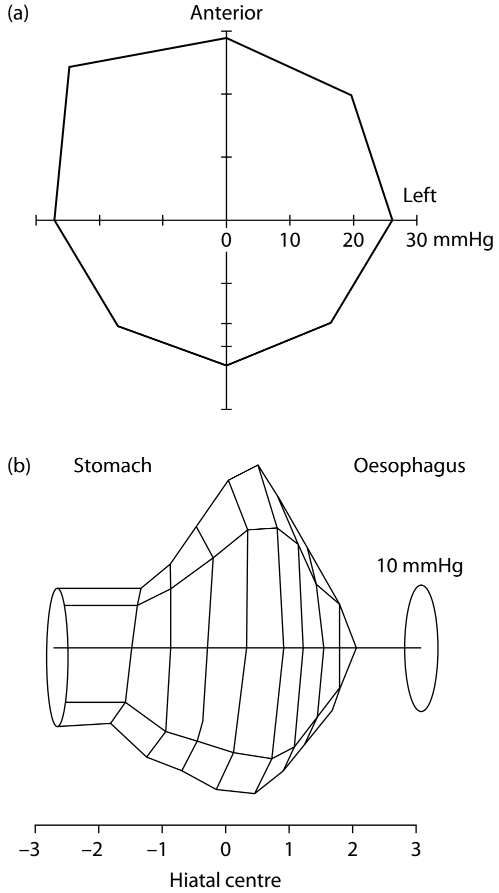


Figure 22.4 Radial (a) and axial (b) pressure profile of the gastro-oesophageal junction in an asymptomatic subject obtained using the technique of vector manometry. Position zero on the axial scale is the mid-point of the diaphragmatic hiatus.

The extrinsic component is largely determined by contractions of the diaphragmatic crura. The increase in LOS pressure is mainly during inspiration and can be as much as 100 mmHg with maximum diaphragmatic contraction. This pinchcock action of the diaphragm is a protective mechanism against reflux induced by sudden increases in intra-abdominal pressure. This mechanism is obviated when a hiatal hernia is present.

Assessment of oesophageal disease

This includes a careful history, physical examination and appropriate investigations to establish the nature of the underlying pathology.

Symptoms

The presentation of oesophageal disease is often typical with one or more of the well-known classic symptoms. Atypical presentation is not, however, infrequent, and oesophageal disease may be mistaken for cardiac and pulmonary disorders. In these patients differentiation is only possible after specialized investigations are carried out. A small cohort of patients with oesophageal symptoms have no abnormality on physical examination and intensive investigations. In some, but not all, of these patients, the symptoms reflect a psychoneurotic state. The typical symptoms of oesophageal disease are dysphagia, regurgitation, odynophagia, chest pain and water brash.

Dysphagia

Difficulty in swallowing or a sensation of food bolus arrest or delay during swallowing may be due to mechanical obstruction or functional disorder. The patient feels the food sticking and often points to a particular site on the sternum, although this does not correlate well with the exact anatomical location of the obstruction. Dysphagia for solids implies significant disease which may be mechanical or functional whereas dysphagia for liquids only is more likely to be functional (oesophageal motility disorder). In the latter, difficulty with swallowing may be intermittent or its severity variable with exacerbations and periods of relative remission. Some patients with dysphagia find that food transit through the oesophagus can be facilitated by sipping fluid after each solid bolus or by repeated swallows and various postural manoeuvres such as expiration against a closed glottis (Valsalva), etc. On the other hand, persistent and progressive dysphagia indicates mechanical narrowing of the oesophageal lumen. This is usually associated with regurgitation and is not relieved by sipping fluids or repeated swallowing. Eventually, with progression to total dysphagia, the patient is unable to swallow saliva and exhibits constant drooling. In obstructive dysphagia, the symptom begins when 20–30% of the oesophageal lumen is lost, and patients usually present when 50% of the oesophageal lumen is compromised.

The sensation of a substernal lump (globus)

When this is present a short period after eating or when fasting it is termed 'globus hystericus'. It is a neurotic symptom in patients with emotional instability but requires thorough examination to exclude organic disease. Dysphagia is never an expression of a purely psychiatric disorder. However, some patients with well-established oesophageal disease may report that their dysphagia is worse during severe emotional periods.

Regurgitation

This symptom results from regurgitation of gastric or oesophageal fluid into the throat accompanied by a sour taste in the mouth. It is often postural and occurs predominantly in the supine position especially at night, with the regurgitated material often staining the pillow. Postural regurgitation, which is a very common symptom of reflux disease, is precipitated by meals and activities associated with a rise in the intra-abdominal pressure, i.e. bending and straining. Regurgitation may also occur as an overflow phenomenon due to the accumulation of food in the oesophagus proximal to a stenosing lesion. This spillback into the pharynx and mouth at night may lead to aspiration pneumonitis. In oesophageal motility disorders both overflow and postural regurgitation may occur, although the former is more commonly encountered in these conditions.

Odynophagia

This complaint consists of localized pain, usually in the lower sternal region, which occurs immediately on swallowing certain foods or liquids. It always indicates organic disease, most commonly oesophagitis. Hot drinks, acid citrus beverages, coffee and heavily spiced foods are among the most frequent dietary items that induce this symptom. It can be severe enough to condition patients not to eat or drink the offending item, or food in general. Odynophagia can be seen after involvement of the mucosa by reflux, radiation, viral or fungal infections. Less commonly, odynophagia can be a manifestation of ulceration or cancer of the oesophagus.

Heartburn

This is the most common manifestation of oesophageal disease and may occur in up to 50% of the population. It is due to reflux of gastric juice, which is injurious to the oesophageal mucosa. The chemical injury is accentuated by a defective clearing of the refluxate by the oesophagus consequent on an impaired motility. This increases the contact time of the acid and any other injurious substance (e.g. bile salts) with the oesophageal mucosa. Some patients complain of severe heartburn, yet on endoscopy there is little or no evidence of inflammation. These individuals may still have reflux with an abnormal oesophageal mucosal sensitivity. Heartburn is often worsened by recumbency, increase in intra-abdominal pressure and may follow fatty meals or alcoholic beverages. Heartburn is usually relieved, even temporarily, by taking antacids. This symptom can increase in intensity until it is perceived as chest pain.

Chest pain

Oesophageal anterior chest pain is often described as a tightening or gripping pain, which closely simulates angina pectoris. Thus it may radiate to the back, jaw, arm and ear and may even be relieved by sublingual nitrates. This type of pain is commonly found in patients with reflux oesophagitis or oesophageal motility disorders. It may occur in association with meals when it persists for about an hour after, but is also experienced in the fasting state and is frequently precipitated by emotion and exercise.

Water brash

This symptom is uncommon and is restricted to patients with reflux disease. It is due to excessive salivation, the mouth becoming full of fluid, which has a salty taste, clear and frothy.

Atypical presentation of oesophageal disease

Patients with oesophageal disease may present with anaemia due to chronic blood loss and, less commonly, with acute upper gastrointestinal bleeding (haematemesis, melaena). Chronic blood loss is usually due to erosive oesophagitis and active bleeding results from the Mallory–Weiss syndrome or peptic ulceration in a hiatus hernia. Incarceration and strangulation of a paraoesophageal hiatus hernia and spontaneous perforation of the oesophagus (Boorhave syndrome) present acutely with a severe life-threatening illness.

Reference has already been made to the frequently encountered difficulty in distinguishing oesophageal from cardiac pain. Often, patients are treated for angina for a while until persistence/aggravation of symptoms indicates the need for coronary angiography. Approximately 20–40% of patients with chest pain and normal coronary angiography are subsequently found to have oesophageal disease.

Presentation with pulmonary symptoms is common. These include attacks of coughing, choking and repeated chest infections due to aspiration pneumonitis in patients with overflow or postural regurgitation. The chest radiograph shows areas of consolidation, abscess formation and pleural effusion. Furthermore, intrinsic asthma is often exacerbated by gastro-oesophageal reflux with aspiration particularly in infants and children. Effective treatment of the reflux disease is often followed by a considerable improvement in the asthmatic condition of these patients.

Physical signs

The oesophagus is a mediastinal structure and is inaccessible to physical examination. However, patients presenting with oesophageal diseases may have physical signs, which should be sought during the examination. These include evidence of weight loss; pallor due to anaemia; swelling in the neck due to pharyngeal pouch; enlarged lymph nodes in the left supraclavicular or cervical regions; chest signs on auscultation and percussion of the lung fields; epigastric mass due to carcinoma of the cardia enlarging downwards, hepatomegaly with or without clinical jaundice; tylosis of the hands and/or feet.

Physiological tests

These include manometry, pH-metry and tests to assess bile reflux. The Bernstein test continues to be used when a correlation between symptoms and acid reflux is required. The standard acid reflux test and the acid clearance test have largely been replaced by pH-metry in the assessment of patients with reflux symptoms.

Oesophageal manometry

This technique measures the mechanical function of the oesophageal musculature and its sphincters by recording intraluminal pressure profiles caused by the contractions. Intraluminal pressure recording is carried out using a system of water-perfused catheters or solid-state strain gauge transducers built into catheters. For sphincter pressure measurements there is a potential for the position of the sphincter to alter in relation to the pressure sensor on the catheter, especially in prolonged measurements. In these situations the Dent sleeve can be used. This is a thin, 6 cm long, open-ended, silastic sleeve which surrounds the pressure sensor ports of a water-perfused catheter. The sleeve operates by traversing the sphincter and records the averaged circumferential and axial pressure forces acting on the sleeve. An alternative is the sphinctometer. This device is usually sited distally in a multichannel microtransducer catheter and consists of a side-mounted transducer surrounded by a silicone, oil-filled silastic tube of 6 cm length and of the same diameter as the catheter. The water-perfused catheter is connected to a low-compliance hydraulic pump, which in turn is connected to a system of strain gauges and a system of readout, most commonly a computer (Figure 22.5). The solid-state strain gauge microtransducers can be connected directly to the readout system. The study commences by determination of the position of the LOS/HPZ. This is done by inserting the catheter into the stomach and withdrawing it either rapidly (rapid pull-through) or slowly (station pull-through) taking pressure recordings at each of the stations from both the sphincter and oesophageal body. The rapid pull-through technique provides information on the position and length of the sphincter. The manometry catheter can then be positioned with one pressure sensor in the middle of the sphincteric area and three pressure sensors separated by 5 cm intervals on the catheter lying within the oesophageal body. In stationary manometry, pressure recordings are taken from the sensors in response to a number of water swallows separated by 30 seconds.

This method provides sufficient information on the position, length and pressure of the LOS in addition to pressure, peristalsis and propagation of contractions within the oesophageal body (Figure 22.6). The information obtained from manometric studies is analysed according to set criteria to diagnose oesophageal motility disorders. This method is indicated for the accurate placement of pH electrodes prior to 24 hour ambulatory pH monitoring in patients suspected of having gastro-oesophageal reflux disease (GORD). This method can also be diagnostic in patients with classical motility disorders of the oesophagus.

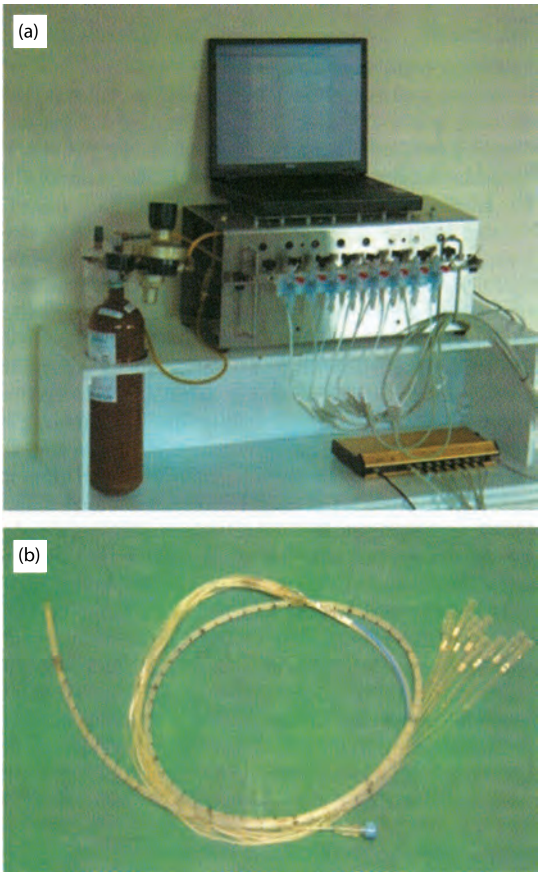


Figure 22.5 Stationary manometry. (a) This is performed using a low-compliance gas-driven hydraulic pump infusing water through a multichannel catheter connected to a system of strain gauges which feed back the pressure signals into a digital converter which is displayed on the computer screen; (b) multichannel silicone catheter used in manometry.

In patients where the stationary manometry reveals a motility disorder of the oesophagus and in those patients suspected of having an oesophageal motility disorder, prolonged ambulatory manometry is preferred. This method employs solid-state pressure transducer catheters attached to a portable recording system with event markers triggered by the patient when symptoms occur (Figure 22.7). They are particularly useful for patients with non-cardiac chest pain who may have transient motility disturbances in the oesophagus and for patients with non-specific motility disorders. Because of the diverse aetiology of these disorders, a combined recording of manometry and pH-metry is indicated.

Analysis of both manometry and pH-metry may reveal the disorder (Figure 22.8). More recently, vector manometry of the LOS has been obtained using a water-perfused catheter with eight radial channels. Although this technique has provided illustrative three-dimensional representation of the LOS, the technique suffers from poor reproducibility.

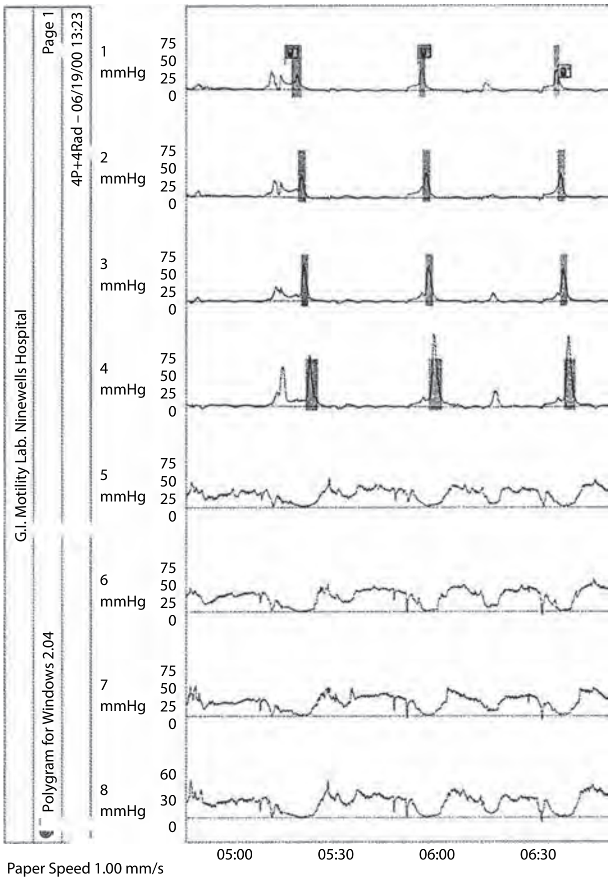


Figure 22.6 Station manometry recording using an eight-channel catheter with four radial channels at the same distance from the tip of the catheter (lower four) positioned within the lower oesophageal sphincter determined by a station pull-through technique and four channels separated by 5 cm intervals. The contractions of the oesophagus and the lower oesophageal sphincter in response to three wet swallows is demonstrated.

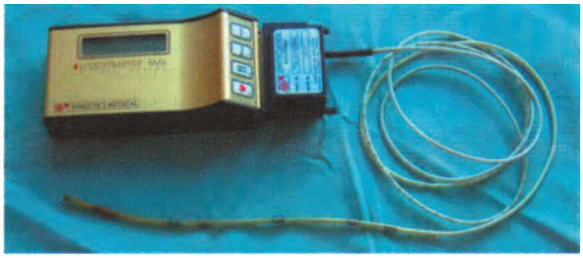


Figure 22.7 A portable recording device connected to a catheter on which three pressure transducers measure pressures and a pH electrode measures the hydrogen ion concentration.

24-hour pH monitoring

This technique involves the transnasal placement of a pH measuring electrode sited 5 cm above the manometrically identified LOS. The pH electrode monitors the changes in intraoesophageal pH over a circadian cycle (24 hours) with the information stored in a portable recording device (Figure 22.9).

A 24-hour pH profile is subsequently obtained from the recording device. This includes the frequency, duration and pattern of reflux episodes together with temporal correlation with symptoms (Figure 22.10). Conventionally, a reflux episode is defined as a pH drop to below pH 4. Specially designed software analyses the data in two ways. In reflux event analysis, all individual reflux episodes are identified and characterized: their number, mean duration, number of long reflux events more than 5 minutes and the duration of the longest reflux episodes. Cumulative oesophageal exposure analysis depicts the frequency distribution of the oesophageal pH data for the erect and supine parts of the study as well as for the whole period of study. The generated data for any one patient are then compared with that which pertains to a 'normal' population. Indications for 24-hour ambulatory pH monitoring include the definitive diagnosis of gastro-oesophageal reflux in patients who are sufficiently symptomatic to have warranted endoscopy, but in whom endoscopy is normal. In addition, it is used to investigate patients suspected of having gastro-oesophageal reflux as a cause of atypical symptoms, such as non-cardiac chest pain, respiratory and laryngeal symptoms, in whom the relevant investigations have been normal and to correlate such symptoms with reflux episodes. Further, it is used when established gastro-oesophageal reflux responds poorly to optimal medical therapy and particularly when surgical treatment is contemplated. It may also be used in the assessment of patients with complex oesophageal disorders prior to surgery.

24-hour oesophageal bile monitoring (by spectrophotometry)

In the last decade, an optical fibre sensor has been developed which can spectrophotometrically detect bilirubin in the oesophagus as a marker of enterogastro-oesophageal reflux. A fibreoptic probe inserted transnasally into the oesophagus and positioned 5 cm above the manometrically determined HPZ is connected to an ambulatory spectrophotometer. The system is marketed under the name 'Bilitec 2000' (Figure 22.11). This device registers the presence of bilirubin when it detects an absorption peak around 453 nm. This peak is said to be easily detectable both in the bile spectrum and in the spectrum of gastric juice combined with other substances (e.g. bile, food, drinks and blood). The presence of bilirubin is, for practical purposes, equivalent to the presence of bile, and the detection of an absorption peak around 453 nm implies the presence of bile. The ambulatory spectrophotometer transfers the signals to a data logger and a 24-hour profile can be recorded. The data can then be offloaded to a computer and an absorbance curve (for 453 nm) is produced for the recording period (24 hours). An episode of bile reflux is defined as an increase in absorbance over 0.14 absorbance units. This threshold has been set arbitrarily to avoid false-positive results, which can be caused by the absorbance at 453 nm of dietary substances consumed during the test. It is important to use a standard diet avoiding coloured food substances and beverages during the test. In addition, bile reflux in an acidic medium ($\text{pH} < 3.5$) may be underestimated as a result of dimer isomerization of bilirubin and a shift in absorption wavelength in this environment. Further, in a few medical conditions (Gilbert and Dubin–Johnson syndromes), there is a disproportionate secretion of bilirubin compared to bile acids. This test is indicated in patients with symptomatic gastro-oesophageal reflux with

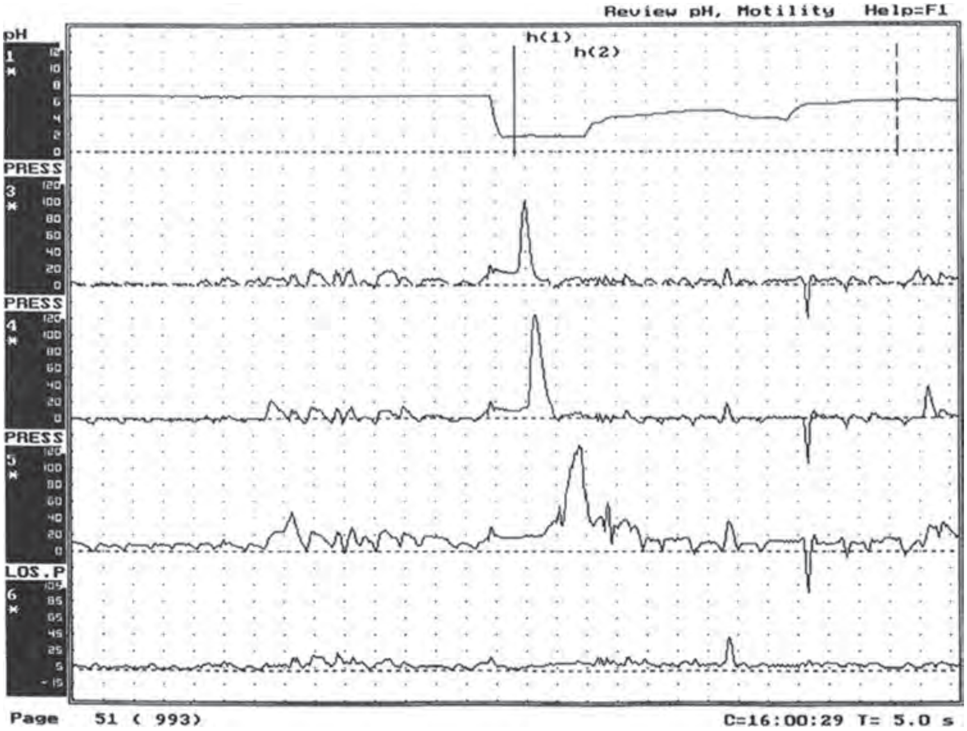


Figure 22.8 Part of a recording of 24-hour combined pressure and pH study showing the drop in pH in the lower oesophagus stimulating a peristaltic contraction to clear the oesophagus of the acid.

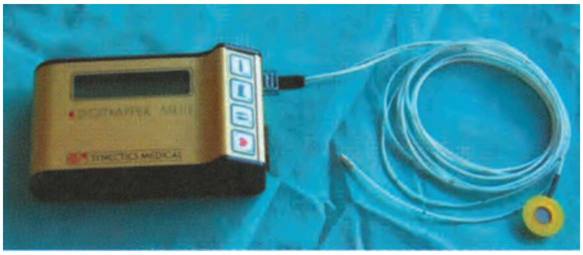


Figure 22.9 Portable recording device connected to a pH catheter with a surface electrode used for 24 hour pH monitoring.

poor response to an adequate dose of proton pump inhibitors (PPIs). It is also indicated in patients with complications of GORD such as Barrett's metaplasia, strictures and ulcers and in patients with reflux symptoms after gastrectomy. Normal values for this test have been published. Validation tests suggest that the overall accuracy of the system is sufficient for clinical use but not reliable enough to study disease mechanisms caused by duodenogastro-oesohageal reflux.

Bernstein's acid perfusion test

This detects oesophageal mucosal sensitivity to acid and is very useful in the determination of the oesophageal origin of chest pain. The patient is studied in the fasting upright position. A nasogastric tube is positioned in the middle of the oesophagus. Infusion is initially started with isotonic saline and then switched to 0.1 M HCl without informing the patient. Acid perfusion will reproduce the pain in patients with a sensitive oesophageal mucosa. After stopping the infusion the pain should subside by 20 minutes.

Motility disorders of the oesophagus

The oesophageal motility disorders are a continuum of functional conditions that affect the oesophagus and are diagnosed by the presence of specific manometric patterns assessing oesophageal peristalsis and LOS pressure and relaxation (Figure 22.12). Symptomatically, they are most frequently associated with dysphagia and chest pain and occasional regurgitative symptoms.

Although several authorities have attempted classification of the variety of abnormal manometric profiles encountered, none has found universal acceptance in the medical community, particularly as the clinical significance of many of these patterns remains unclear. More frequently, oesophageal motility disorders are classified as being of primary and idiopathic (where the oesophagus is the originating site of pathology but its exact cause is unknown) or secondary origin (oesophageal dysmotility being the result of a systemic condition) (Box 22.1).

This section will focus on those oesophageal motility disorders where characteristic radiological and manometric features have allowed for a well-formed clinical definition and the development of medical and surgical treatment modalities targeting both patient symptomatology and the pathophysiological process as it is currently understood.

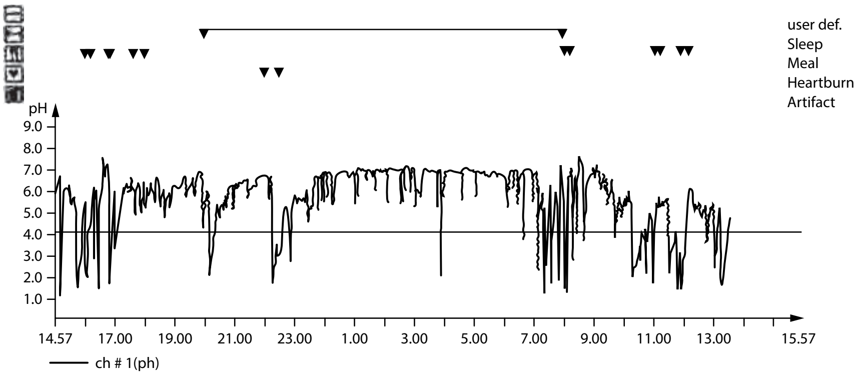
Achalasia

Epidemiology

Achalasia (lit. failure to relax) is the most regularly diagnosed of the primary motility disorders with an annual incidence of 0.4–0.6/100 000 and point prevalence of 8/100 000 people.

pH Channel

Compressed 24-hour pH graph



Period Table

Item		Total	Upright	Supine	Meal	PostP	HrtBrn
Duration of period	(HH:MM)	21:20	10:35	12:00	01:15	07:22	00:30
Number of acid refluxes	(#)	33	26	12	8	14	3
Number of long acid refluxes	(#)	9	6	3	1	4	2
Longest acid reflux	(min)	26	26	8	8	16	8
Total time pH below 4.00	(min)	144	124	36	16	58	14
Fraction time pH below 4.00	(%)	11.2	19.6	5.0	21.6	13.0	45.2
Symptom index	(%)	n/a	n/a	n/a	n/a	n/a	100.0

Figure 22.10 A compressed recording of 24-hour pH-metry showing episodes of acid reflux in the lower oesophagus with analysis of the reflux episodes.

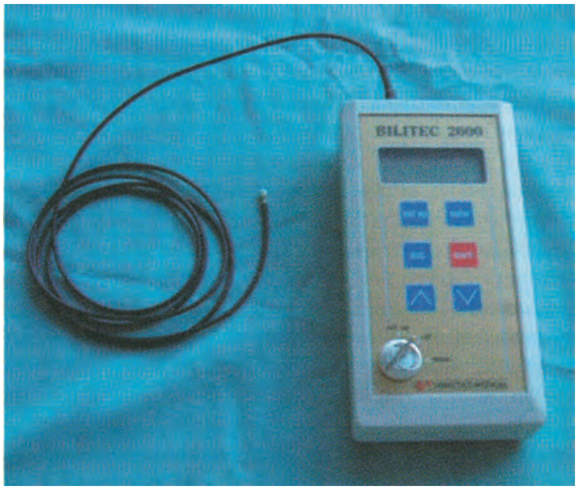


Figure 22.11 An ambulatory spectrophotometer connected to a fibreoptic probe used for prolonged ambulatory bile reflux monitoring.

Pathology

To date, no single aetiological factor has been identified as being causative of achalasia. However, histological analyses of oesophageal samples obtained at autopsy and during surgery have consistently demonstrated pathological changes in the myenteric (or Auerbach's) plexus. This network of nerves, a part of the enteric nervous system, is located between the longitudinal and circular muscles in the muscularis externa and provides motor innervation along the gut. The typical features seen in the plexus are a patchy inflammatory response with infiltration of T-lymphocytes, eosinophils and mast cells which eventually leads to neural fibrosis and permanent destruction of the postganglionic inhibitory neurones. These inhibitory neurones are responsible for the release of the neurotransmitters NO and VIP, which dampen the effects of postganglionic cholinergic neurones and result in relaxation of the LOS. As a result, there is an unopposed action of the cholinergic neurones, resulting in a permanently contracted lower oesophagus, one of two principal features of achalasia. Aperistalsis, the second cardinal sign of the condition, results from loss of the progressive neurotransmitter-mediated gradient that permits sequential contractions along the oesophagus.

Clinical features

Many symptoms are associated with achalasia, of which dysphagia and chest pain are the most prominent and frequently reported (Box 22.2). Dysphagia is the most common presenting symptom of achalasia with as many as 94% of patients reporting it in one study. It is usually gradual in onset and present in many cases for many years prior to presentation. The dysphagia will initially be intermittent and to solids only, progressively extending to fluids and is occasionally associated with regurgitation, particularly in the advanced stages when the oesophagus is dilated, and on lying flat, giving a sensation of vomiting or choking.

Chest pain and odynophagia are usually features of early achalasia when oesophageal dilatation is minimal and diminishes as the degree of widening increases. However, as mentioned, regurgitation also increases with dilatation, and some 90% of patients will produce a foamy mucoid saliva. Food stasis and carbonated drinks, which patients consume to empty the oesophagus, are speculated to ferment and produce dyspepsia (in up to 52% patients) and halitosis.

As might be expected, in very advanced achalasia the lack of peristalsis and oesophageal emptying can produce a massively dilated and tortuous oesophagus together with similarly severe

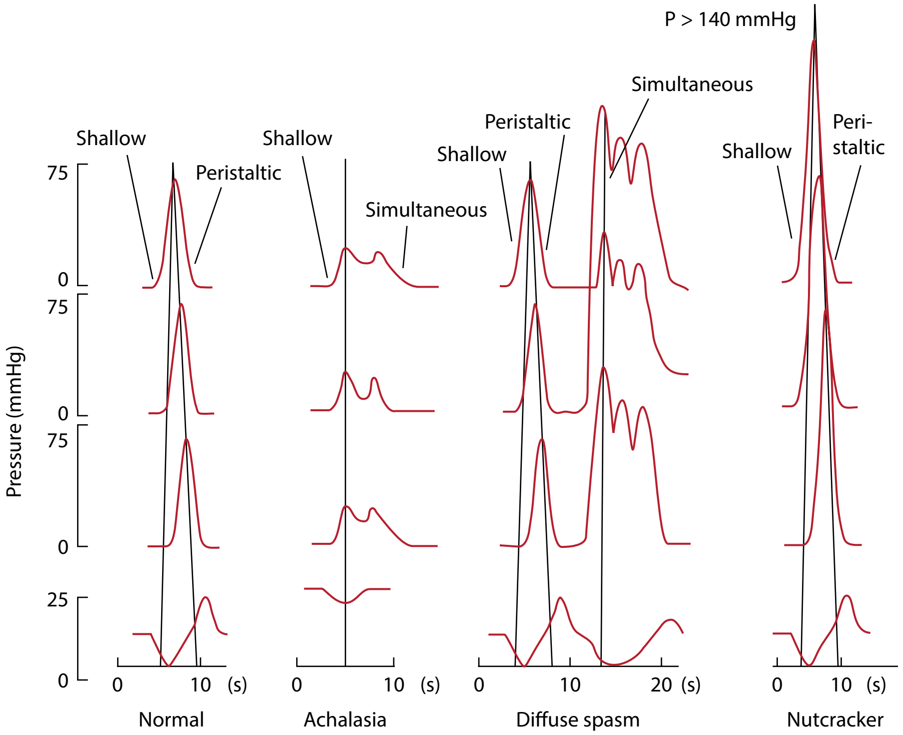


Figure 22.12 Schematic diagram summarizing the normal and disordered motility profiles.

BOX 22.1 Primary and secondary oesophageal motility disorders (note that this is not a comprehensive list of secondary disorders)

Primary causes of oesophageal motility disorders

- Achalasia
- Diffuse oesophageal spasm
- Hypercontracting oesophagus (hypertensive/nutcracker oesophagus)
- Hypocontracting oesophagus (ineffective oesophageal motility)

Secondary causes of oesophageal motility disorders

- Collagen vascular disease (e.g. systemic sclerosis)
- Diabetes mellitus
- Chagas disease
- Amyloidosis
- Alcohol excess
- Myxoedema
- Multiple sclerosis
- Idiopathic pseudo-obstruction
- Old age
- Dermatomyositis
- Systemic lupus erythematosus
- Malignancy

dysphagia, regurgitation, weight loss (as food is undigested), anaemia and respiratory complications, e.g. pneumonia.

As 3–7% of patients with achalasia develop cancer, it is important to have a high suspicious rate of cancer if patients develop new onset of symptoms. A repeat endoscopy is advisable in those patients.

BOX 22.2 Clinical features of achalasia

- Dysphagia
- Chest pain
- Regurgitation
- Odynophagia
- Dyspepsia
- Halitosis
- Weight loss
- Mega-oesophagus
- Anaemia
- Respiratory complications, e.g. pneumonia

Diagnosis

Although there are many symptoms that can be indicative of achalasia, as can be readily appreciated, they are very non-specific and consequently a more objective assessment is needed to confirm the diagnosis.

Radiology

Barium swallow with fluoroscopy has traditionally been the first-line investigation for suspected achalasia. However, with the advent of manometry, its role is diminishing, but remains important as a relatively cheaper and more easily accessible imaging modality.

In the early part of the disease, the study will usually show a normal diameter oesophagus, with some loss of normal peristalsis. Progressively, the disease will result in an increasingly dilated and tortuous oesophagus with retained food and saliva being demonstrated as an air–fluid interface. The classical features, however, will be of a dilated distal oesophagus which gradually

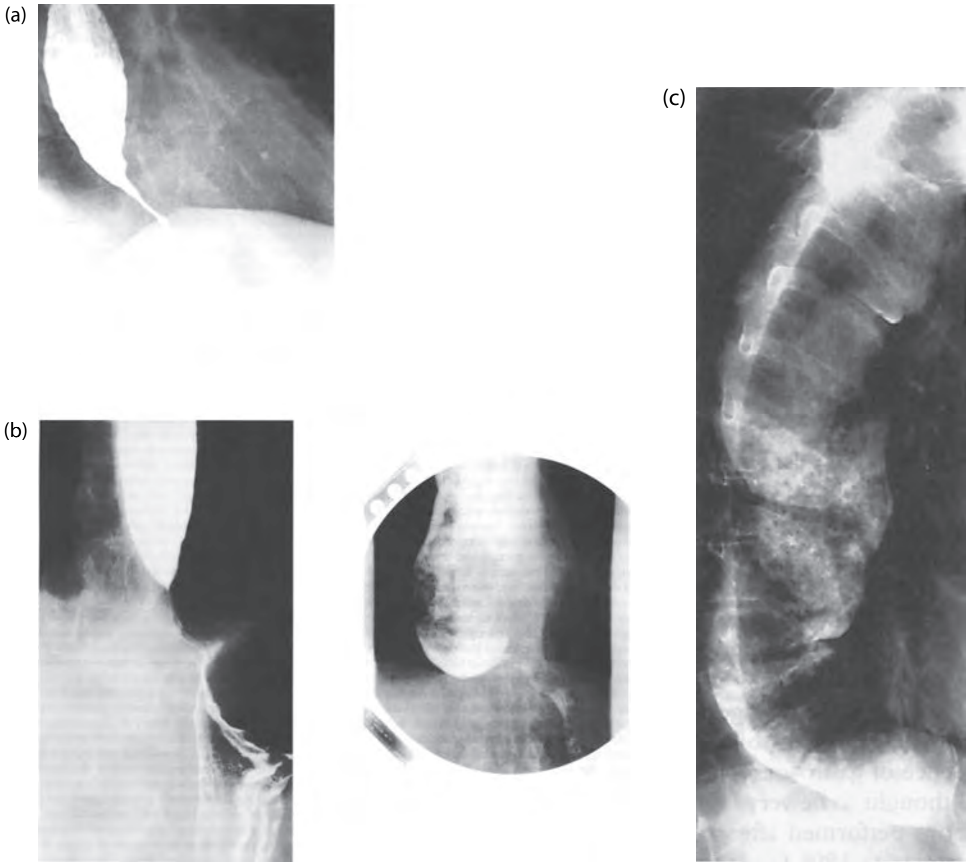


Figure 22.13 (a) Barium swallow in a patient with achalasia of the oesophagus is dilated above an apparent smooth narrowing of the cardia. The tapering of the oesophagus to the cardia has been likened to a 'bird beak'; (b) barium swallow showing grossly dilated sigmoid oesophagus in a patient with longstanding achalasia; (c) barium swallow in longstanding achalasia. The patients presented with total dysphagia and a neck lump which disappeared after insertion of a nasogastric tube. The grossly dilated and lengthened oesophagus has assumed a sigmoid shape. These patients do not benefit from cardiomyotomy and, if fit, are best treated by oesophagectomy with gastric tube replacement.

tapers at its lowermost portion giving an appearance of a 'bird's beak' (Figure 22.13).

Plain posteroanterior chest radiographs although of no significant diagnostic use can occasionally also show a soft-tissue shadow with a fluid level and an absent fundic air bubble in some 50% of cases and may be an indicator for further investigations in the undiagnosed individual.

Manometry

In the early stages of the disease, achalasia may not always show up on a swallow study due to the insufficient dilatation of the oesophagus. However, manometry is able to identify even the very earliest cases convincingly and confirm the diagnosis without significant delay. It is therefore increasingly becoming the first-line investigation when available (Figure 22.14).

Lower oesophageal, smooth muscle manometry is the most useful of all the available investigative techniques and regarded as the gold standard when diagnosing oesophageal motility disorders. Many investigators have propagated classification systems for dysmotility syndromes based on these unique manometric findings, but the findings have generally been too diffuse to produce any consistent, acceptable methodology.

However, in achalasia there is a clear known manometric pattern:

- Aperistalsis of the oesophagus is present and wet swallows are immediately followed by identical simultaneous contractions at an amplitude of 10–40 mmHg.
- There is an abnormal relaxation mechanism of the LOS associated with achalasia. Around 80% of patients will have either completely absent or incomplete relaxation on swallowing food or fluid, whereas manometry in the remaining 20%, which may exhibit complete LOS relaxation, is associated with short duration (<6 seconds) relaxations, which are functionally insufficient to completely empty the lower oesophagus.

Manometry has also led to the definition of a unique condition termed '*vigorous*' achalasia. This manometric entity has features of both classical achalasia and diffuse oesophageal spasm (DOS; see below), characterized by repetitive, high (or normal) amplitude aperistaltic oesophageal contractions occurring in concert with a non-relaxing LOS, where again the pressure might be high (>45 mmHg) in some 50% of patients. Clinically, it can be associated with a higher degree of chest pain than its classical counterpart but the remaining symptoms will be the same. Its management also differs very little from the standard therapy but surgical treatment, if indicated, will involve a longer myotomy extending from the level of the aortic arch to the GOJ.

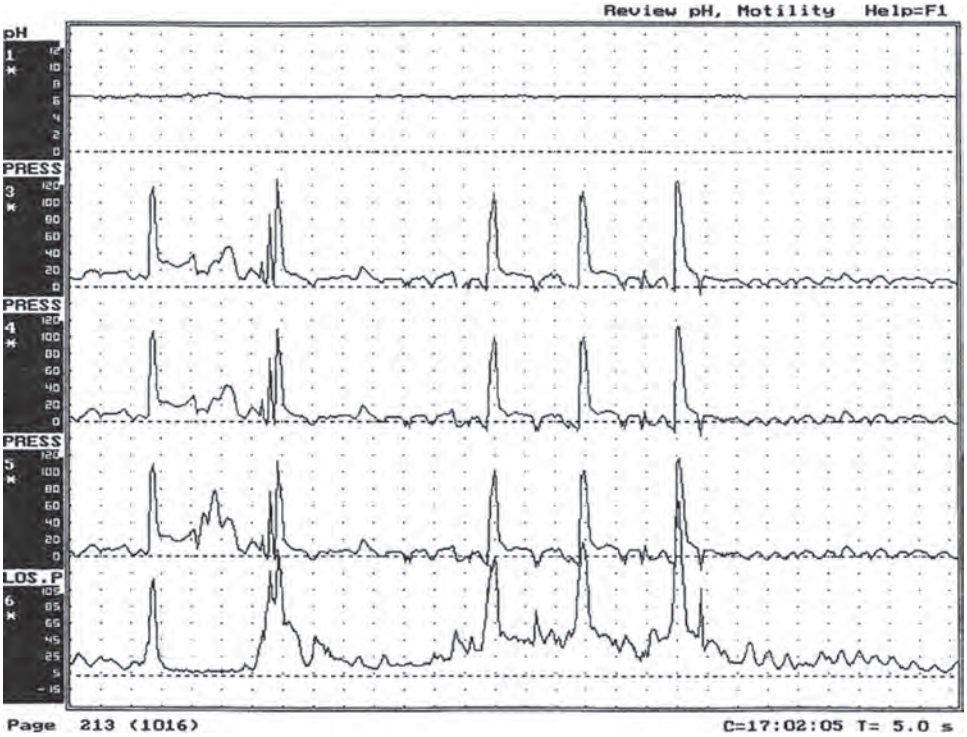


Figure 22.14 Manometric tracing obtained from a 24-hour combined pH and manometry study on a patient with achalasia. The top channel shows the pH in the lower oesophagus. Channels 3, 4 and 5 (5 cm apart in the body of the oesophagus) show aperistaltic contractions in the oesophagus, and channel 6 (in the lower oesophageal sphincter) shows failure of relaxation.

Endoscopy

Endoscopy, alongside endoscopic ultrasound (EUS) and CT, is not regarded as a direct diagnostic tool for achalasia. However, it is vitally important to exclude any organic pathology such as oesophagitis, reflux and gastro-oesophageal malignancy which may mimic its symptomology. Consequently, oesophagogastric duodenoscopy (OGD) should be carried out in patients presenting with dysphagia and chest pain. In the hands of an experienced endoscopist, evidence of reduced peristalsis and food stasis may be evident in advanced achalasia and again be an indication for further definitive investigation with barium swallow studies or manometry.

Management of achalasia

Several treatment modalities have been developed over the past century to treat achalasia, although none is consistently able to restore oesophageal peristalsis and normal relaxation of the LOS. Medical and surgical treatments are therefore used for symptomatic control by reducing the pressure gradient across the LOS, thereby reducing stasis in the lower oesophagus and its associated complications.

Medical treatment

Pharmacological treatment methods have traditionally been reserved for frail, elderly patients, those who have significant comorbidities or individuals who are unwilling to undergo surgical intervention, which appears to achieve the most sustained symptomatic control in affected patients.

Calcium channel blockers and long-acting nitrates have been used to reduce LOS spasm. Long-acting nitrates in particular are able to partially antagonize the effects of the unopposed postganglionic cholinergic fibres and have been shown to reduce LOS pressure by 66% for up to 90 minutes at a time. As a result it is often taken prophylactically or before meals to aid oesophageal emptying. An initial response is seen in as many as 70% of patients; however, its efficacy diminishes over time. A recent Cochrane Library analysis has also reported that there is insufficient evidence to give suitable implications for its off-licence use in achalasia.

Endoscopic injection of botulinum

Endoscopic injection of botulinum toxin A into the LOS has also been vigorously advocated in recent years as an effective treatment that can be readily repeated as required. The toxin, produced by a member of the *Clostridium* family, inhibits acetylcholine release at the postganglionic nerve terminals (which is responsible for the increased LOS tone), thereby, in essence, replacing the natural inhibitory function of the damaged postganglionic NO and VIP neurones. Botulinum is effective for short-term symptom control in 85–90% of patients but symptoms tend to recur within 6 months in up to 50% of cases. As with oral agents, the effectiveness of botulinum tends to reduce with repeated injections.

Both oral therapy and endoscopic injection of botulinum toxin, although useful temporizing measures for symptom relief in the short term, remain inferior to pneumatic balloon dilatation and myotomy as more definitive treatment options. Furthermore, a recent Cochrane review confirmed that pneumatic balloon dilatation offers greater symptom control over a period of 6–12 months than botulinum injection.

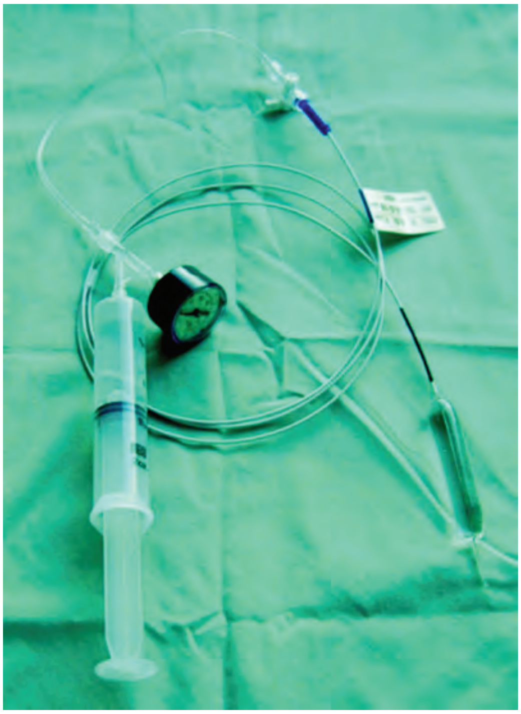


Figure 22.15 Hydrostatic polyurethane balloon dilator.

Pneumatic dilatation

Pneumatic dilatation has been used in clinical practice for many decades and involves endoscopically employing a deflated 30–40 mm diameter balloon across the LOS (Figure 22.15). This can be carried out via endoscopy or under fluoroscopy. On inflation of the balloon, which is usually performed no more than twice in the same sitting, the muscle fibres of the LOS are spread apart and torn. Patients tend to recover very quickly from the procedure, which can be performed in the outpatient setting, with studies showing symptomatic relief in 60–93% of patients, with the response being proportional to the diameter of the balloon deployed. Simultaneous fluoroscopy is frequently employed at the time of dilatation to show the passage of contrast (and food) into the stomach, indicating the success of the procedure.

Although its symptomatic control is excellent, the procedure is associated with a 2% risk of oesophageal perforation and gastro-oesophageal reflux. Additionally, some 30% of patients may require a further dilatation and around half will develop symptom recurrence over a 5 year period.

From a surgical perspective, balloon dilatation (together with botulinum infiltration) may result in scarring and destruction of the surgical planes at the GOJ and therefore may make subsequent cardiomyotomy more difficult and prone to a more uncertain outcome. However, in expert hands the effect of previous balloon dilatation or botulinum infiltration on surgery is minimal.

Surgical treatment

Laparoscopic cardiomyotomy

The standard surgical management of achalasia is laparoscopic cardiomyotomy. The myotomy that was described by Heller in 1913 involved performing two myotomies, one anteriorly and the other posteriorly, on the oesophagus. The procedure was refined by Zaaijer in the 1920s with a single myotomy being made only on the anterior surface. The current standard clinical practice is performing myotomy of the muscle coat of the distal oesophagus up to the dilated segment as well as 2–3 cm of proximal stomach. The vagus nerve should be identified and protected (Figure 22.16). Great care should be taken to avoid injury to the mucosa. Although several methods are used to divide the muscle layer to expose the mucosa, the authors' preference is distraction of the muscle using graspers without any use of energy devices such as diathermy or harmonic dissectors in order to avoid late diathermy injury. Some surgeons use a balloon to distend the GOJ and facilitate division of the muscles. Modern practice may also involve performing a partial fundoplication to prevent postoperative gastro-oesophageal reflux (Figure 22.17).

Overall surgical approaches to achalasia have managed to achieve good outcomes in 80–100% of patients with poor results being usually attributed to poor surgical technique (incomplete myotomy) or development of gastro-oesophageal reflux (due to insufficient fundoplication). The procedure has an associated morbidity of approximately 0.1%.

Although, surgery has been widely accepted as the gold standard intervention, patients are still carefully selected usually on the following criteria:

- symptom recurrence despite endoscopic therapy (botulinum infiltration and balloon dilatation)
- severe symptoms
- children in whom morbidity is increased by repeated procedures.

Several areas of debate have arisen over the years as to the best techniques:

- *Laparoscopy vs thoracoscopy.* The first laparoscopic Heller's myotomy was reported in 1991. As in the open era, the thoracoscopic approach

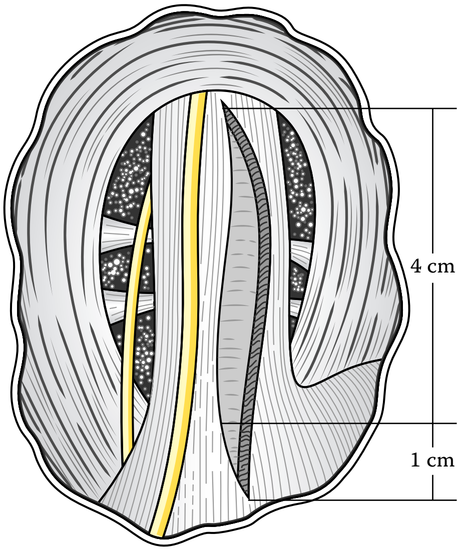


Figure 22.16 Schematic representation of modern cardiomyotomy with protection of the anterior vagus nerve.

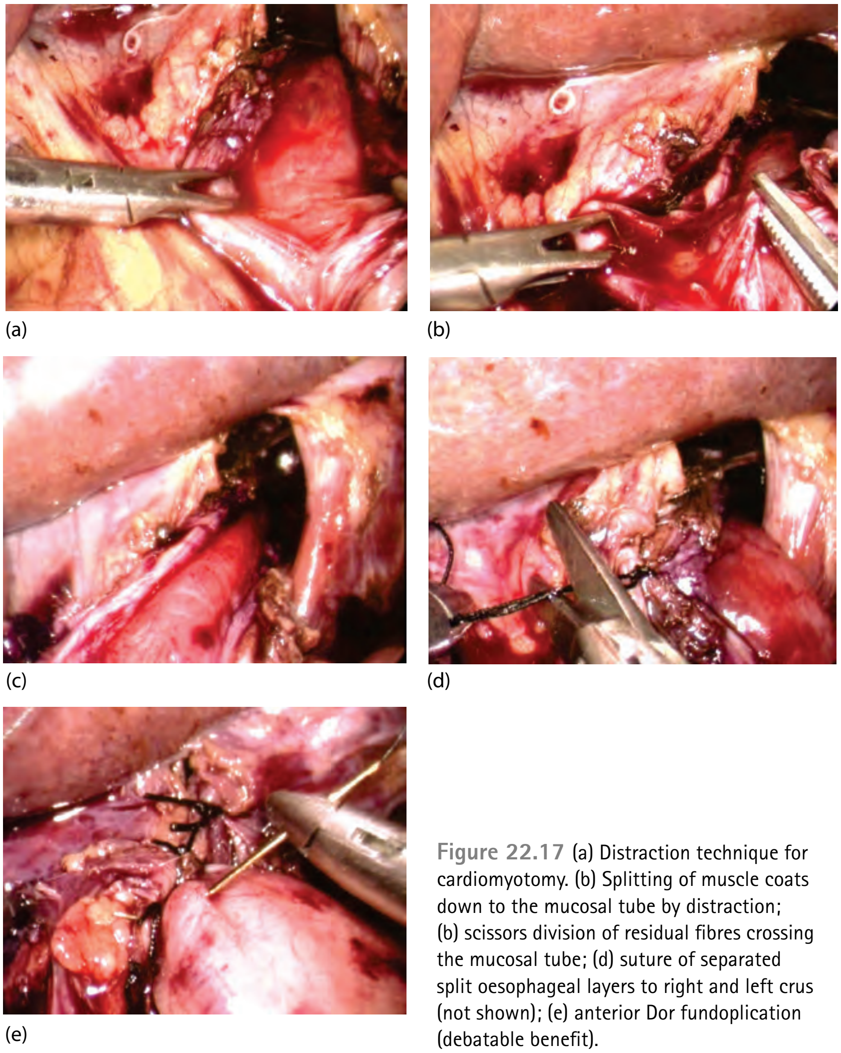

(a) (b) (c) (d) (e)

Figure 22.17 (a) Distraction technique for cardiomyotomy. (b) Splitting of muscle coats down to the mucosal tube by distraction; (b) scissors division of residual fibres crossing the mucosal tube; (d) suture of separated split oesophageal layers to right and left crus (not shown); (e) anterior Dor fundoplication (debatable benefit).

was also initially popular due to the access it afforded to the oesophagus. However, as in the open era, the same problems with complex anaesthesia and high rates of reflux ensued. Around the same time, it was being increasingly recognized that the myotomy should extend some 2–2.5 cm into the cardia to achieve the best results and that a fundoplication procedure aided in substantially reducing reflux (from 48% with myotomy alone to 9% when combined with a posterior fundoplication). Access to the stomach was easier via laparoscopy once again.

- *Complete (Nissen's) vs incomplete (Toupet/Dor) fundoplication.* In cases of gastro-oesophageal reflux, a 360° Nissen-type fundoplication is favoured, which mimics the action of the LOS to control symptoms. In achalasia, where peristalsis is absent and the LOS contracted, a total fundoplication can further impair the LOS and result in persistent dysphagia. One long-term study found that 82% of patients who had undergone myotomy with a Nissen's fundoplication developed recurrent dysphagia over a 10 year period. Additionally, a study comparing an anterior partial fundoplication (Door) to Nissen's fundoplication with myotomy showed dysphagia occurring in 2.8% and 15% of patients respectively over a 5 year period.
- *Anterior partial (Door) vs posterior partial (Toupet) fundoplication.* At present no randomized trials have reported on the superiority of the two types of partial fundoplication in aiding symptomatic relief. At the time of writing, a prospective European, multicentre randomized controlled trial comparing the two procedures is ongoing, examining postoperative pH monitoring evidence of gastro-oesophageal reflux. Proponents of the posterior approach argue that the myotomy is kept separated whereas those advocating the anterior approach contend that the mucosa can be covered to prevent reflux and there is no need for a posterior dissection, making the procedure faster. Many studies have shown a >90% improvement in dysphagia via the anterior approach.
- *Laparoscopic cardiomyotomy vs balloon dilatation.* There has been a renewed interest in re-examining the efficacy of balloon dilatation compared with the best surgical treatment. A multicentre European trial reported that, in 201 patients who had undergone pneumatic dilatation of laparoscopic Heller's myotomy with a Dor fundoplication, therapeutic success of both procedures at 1 and 2 years was equivalent at approximately 90% with no difference of pressure at the LOS after 2 years. This study would suggest that, although surgery has an important role in achalasia treatment, more marginal surgical patients may benefit equally from dilatation therapy.

Recently several groups have reported initial results with single incision, robotic and per oesophageal myotomy techniques. Although the results appear promising in relation to symptomatic relief and reduction of intraoperative oesophageal perforation, there is at the moment scant evidence to suggest whether these techniques will surpass the current open and laparoscopic approaches.

Oesophagectomy

Subtotal oesophagectomy with anastomosis in the neck is sometimes performed for selected indications such as (1) end-stage achalasia and mega-oesophagus as cardiomyotomy might not be sufficient; and (2) in patients with recurrent achalasia after redo surgery. Non-functioning oesophagus with severe symptoms, inability to maintain weight and recurrent chest infections are the indications for oesophagectomy in those cases. Minimally invasive or transhiatal oesophagectomy are the approaches that are frequently used.

Diffuse oesophageal spasm

DOS is the presence on manometry of simultaneous, uncoordinated contractions of the distal oesophagus in individuals with an otherwise normal peristalsis amplitude and duration and normally responsive LOS.

Epidemiology and pathology

This pattern is found in around 3–5% of individuals undergoing manometry for investigation of suspected oesophageal motility disorders, with patients generally being older than 50 years.

Neural dysfunction has been suggested as a possible cause in DOS, as patients have been found to be hypersensitive to the hormone pentagastrin and cholinergic stimulations. Emotional distress and olfactory stimuli have additionally been implicated in the development of abnormal oesophageal contractions, suggesting a more functional aetiology as the driver for DOS.

Clinical features

Patients are frequently found to have emotional liability with recurrent food- and exercise-induced chest pain being the most reported symptom. Confusingly for the patient and investigating physician, the pain can be very similar in nature to cardiac chest pain (i.e. substernal/left sided, constrictive with radiation to the back, jaw) and may even be responsive to nitrates. Patients will often present to the surgical clinic having had extensive cardiac investigations and a normal coronary angiogram. The more classical oesophageal dysmotility symptoms of dysphagia and reflux are usually much less severe, intermittent and non-progressive. Dysphagia can arise as a result of stress, cold fluid intake, rapid eating and extreme cold/warmth.

Female patients in particular may give a history of irritable bowel syndrome. Interestingly, patients with DOS may progress to display clinical and manometric features of achalasia, which has led some investigators to suggest some overlap, and indeed continuity, between the conditions.

Diagnosis

Owing to the considerable clinical overlap with GORD and angina combined with the advancing age of patients, it is important to investigate patients with a cardiac work-up and consider an OGD to further exclude any organic pathology at the GOJ. Additionally, a trial period of high-dose PPIs may prove useful, especially where the diagnosis remains unclear as some 80% of patients with gastro-oesophageal reflux-related chest pain will report an improvement, whereas only 20% of patients having chest pain without gastro-oesophageal reflux experience symptomatic relief with PPIs.

Manometry

Manometry will demonstrate normal peristaltic contractions, interrupted by simultaneous contractions in more than 20% but less than 100% of wet swallows (the presence of continuous simultaneous contractions is diagnostic of achalasia). The LOS is grossly unaffected

Further tests

Radiological assessment is highly variable and does not seem to be of any specific diagnostic value. However, an ambulatory 24-hour pH monitoring study is useful for picking up cases of GORD, which can concurrently occur in up to 50% of patients.

Barium swallow may demonstrate repetitive tertiary contractions which cause the typical, diagnostic corkscrew appearance. However, this is found in only in 30% of patients (Figure 22.18).

Treatment

If underlying organic pathology such as angina and malignancy has been excluded, often simple reassurance that patient symptoms are benign and not life threatening will be sufficient treatment alone. Furthermore, it is important to treat comorbidities

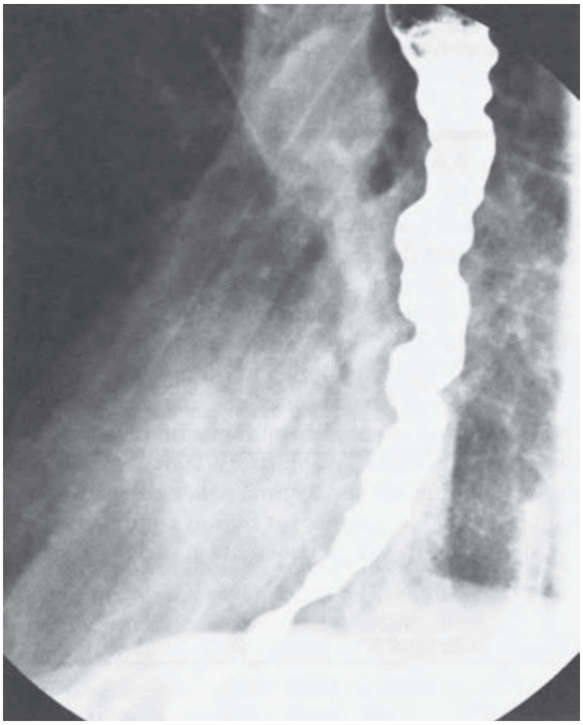


Figure 22.18 Barium swallow in a patient with diffuse oesophageal spasm showing 'corkscrew' appearance but this is only found in 30% of patients.

such as GORD which might be coincidentally identified over the course of the investigation process.

Lifestyle interventions

- Avoid spasm triggers, e.g. dietary modification to soft foods and liquids and stress reduction.
- Reduce smoking and alcohol intake as both are associated with an increase in acid reflux and subsequent oesophageal spasm.

Oral therapy

- Nitrates, e.g. glyceryl trinitrate, isosorbide dinitrate, reduce oesophageal contractile pressure and promote normal peristalsis.
- Calcium channel blockers, e.g. nifedipine, diltiazem.
- Anticholinergics, e.g. atropine, hyoscine.
- Phosphodiesterase type 5 inhibitors, e.g. sildenafil, result in increased local NO concentration in the neuronal terminals by inhibiting breakdown but can be associated with cardiovascular complications.
- Antidepressants, e.g. trazodone, imipramine, reduce patient discomfort but without significantly impacting the pathology. They reduce chest pain in the short term (1–3 months) in 50% of patients.

All the oral agents can reduce high-amplitude oesophageal contractions but are not always effective for symptomatic relief of chest pain. Additionally, none of the above medications is specifically licensed for the treatment of DOS in some countries.

Endoscopic injection of botulinum

Botulinum toxin (see above) is injected at multiple sites along the oesophagus and within the LOS itself. It has been found to completely resolve symptoms in around 50% of patients with a mean duration of 7 months.

Surgical treatment

In patients with intractable symptoms, which are unresponsive to medical therapy, a long surgical myotomy, extending from the level of the aortic arch to the GOJ, is the preferred intervention. However, the results are extremely variable and frequently not useful in improving the patient's symptoms. Thoracoscopic long myotomy may improve chest pain but not dysphagia.

Hypercontracting oesophagus

Hypercontracting oesophagus is a term used to describe individuals who have manometric oesophageal contraction pressures that are two standard deviations above the mean. These patients are further divided into two subgroups.

Hypertensive (nutcracker) oesophagus

The hypertensive oesophagus describes a situation when extremely high-amplitude contractions, greater than 180 mmHg, occur either throughout the oesophagus or in its distal portion. The term is also used when duration of distal peristalsis is >6 seconds, which will subsequently also generate persistently high pressures in the lower oesophagus. Apart from the contractions that are of increased pressure, all other contractions are seemingly unaffected, suggesting that the condition is intermittent.

Hypertensive lower oesophageal sphincter

The LOS is said to be hypertensive when its resting pressure exceeds 45 mmHg; however, peristalsis is usually unaffected.

Interestingly, manometry can often indicate the duality of the two conditions, which has led to the suggestion that they jointly represent a syndrome of hypercontractile oesophagus rather than two distinct entities. As is common with all primary oesophageal motility disorders, to date no aetiological agent has been identified and patients usually have the non-specific symptoms of chest pain (occurs in up to 48% of patients) and dysphagia.

Treatment

Treatment is along the same lines as DOS but the results can be unpredictable. Weak sedatives, antidepressants, behavioural therapy with biofeedback have also been employed to varying levels of success.

Hypocontracting oesophagus (non-specific oesophageal motility disorders)

This group of disorders describes those patients who present with a wide variety of abnormal manometry patterns that do not conform to the classical primary motility disorders but are nonetheless associated with debilitating symptoms. The non-specific motility disorders can broadly be divided into two subgroups.

Ineffective oesophageal motility

Many patients who have non-specific oesophageal motility disorders will have manometry tracings that demonstrate:

- low-amplitude (<30 mmHg) peristaltic contractions (real-time contrast studies of the oesophagus have demonstrated that an amplitude of 30 mmHg is needed to clear the oesophagus)
- simultaneous contractions of the distal oesophagus
- incomplete peristalsis (where propagated peristalsis waves do not cross the whole length of the oesophagus).

These physiological states prevent an adequate clearance of food and debris into the stomach, and consequently patients are termed to have ineffective oesophageal motility. Dyspepsia and heartburn are more common symptoms and patients are frequently noted to have gastro-oesophageal reflux and ENT complaints.

Hypotensive lower oesophageal sphincter

This patient group has a diminished LOS pressure (<10 mmHg; normal, 10–45 mmHg), which is often associated with gastro-oesophageal reflux. This would suggest that the hypocontractility of the oesophagus is the result of reflux-induced acidic damage at the distal oesophagus.

At present, no medication is available to increase peristaltic amplitude. Cisapride (a serotonin receptor agonist acting as a prokinetic agent) was previously used but withdrawn in 2000 due to a significant association with cardiac arrhythmias. Other prokinetic agents such as metoclopramide and domperidone may encourage oesophageal emptying whereas PPIs may help to inhibit the acid reflux and resultant damage at the LOS.

Motility disorders secondary to systemic disease

Secondary motility disorders are the result of a systemic illness causing oesophageal symptoms that mimic those of the primary dysmotility disorders (Box 22.3). Patients are however generally older, with a shorter duration of symptoms and will normally have signs indicating systemic illness.

It is vitally important that patients are thoroughly investigated to exclude not only a primary motility disorder but, more importantly, a malignant condition, such as GOJ tumour, which can often give symptoms similar to achalasia. Treatment will be dependent on the underlying abnormality and findings of manometry but some patients who are otherwise healthy may benefit from pneumatic balloon dilatation.

Although a variety of treatment options are available to treat various conditions, it is important to recognize that oesophageal motility disorders remain a relatively less understood and underappreciated area of surgical practice whose distressing symptoms can result in significant malnutrition, social disability and chronic morbidity. Patients should be investigated thoroughly and treated promptly. It is important to remember that interventions aim to improve the quality of life of those patients and therefore surgery should not be taken lightly if the risks are high and outweigh the benefits desired.

Gastro-oesophageal reflux disease

Introduction

GORD is a chronic, relapsing condition and is one of the commonest reasons for attending primary care. There is a clinical spectrum of disease severity, encompassing patients with mild symptoms to those who require surgical intervention to gain symptomatic control. In the developed world, an approximate prevalence of 10–20% for GORD (defined by at least weekly heartburn and/or acid regurgitation) has been identified whereas in Asia this is lower (at <5%). The incidence in the developed world is estimated to be approximately 5 per 1000 person years. GORD is a significant disease burden and the health service costs associated with GORD are considerable. An estimated four or five per 10 000 people aged 18–60 years take maintenance PPIs for oesophagitis and/or reflux, with the annual drug budget for PPIs in the NHS (UK) in excess of £300 million.

BOX 22.3 Some of the systemic diseases that can present with oesophageal dysmotility (the secondary oesophageal motility disorders)

- Collagen vascular disease (e.g. systemic sclerosis)
- Diabetes mellitus
- Chagas disease
- Amyloidosis
- Alcohol excess
- Myxoedema
- Multiple sclerosis
- Idiopathic pseudo-obstruction
- Old age
- Dermatomyositis
- Systemic lupus erythematosus
- Malignancy

Aetiology of GORD

The aetiology of GORD is complex and multifactorial in nature. Several studies have shown that there is no association between sex and GORD (despite there being a significant association between pregnancy and GORD). The effect of increasing age on the prevalence of GORD is unclear, with two European studies showing a slight but significant association. However, both the UK GP database study and Georgia Medicaid study showed an increase in the incidence of GORD symptoms up to the ages of 69 years and 55 years, respectively, after which time the trend reversed. Increasing prevalence of GORD is associated with higher body mass index, and the Georgia Medicaid study reported a clear positive relationship between GORD diagnosis and obesity [odds ratio 2.8 (95% confidence interval 2.1–3.6)]. Population-based studies comparing reflux symptoms in monozygotic and dizygotic twins have shown that genetic factors might contribute by 31% to the aetiology of symptomatic GORD. Although demographic and genetic factors may have a role in the development of GORD, reflux is likely to be mainly caused by environmental factors. Several behavioural factors are thought to trigger gastro-oesophageal reflux episodes and these include:

- tobacco smoking
- increased salt intake
- excess alcohol consumption
- coffee consumption.

An increased incidence of GORD has also been demonstrated with the use of certain prescription medications. The current use of aspirin, non-steroidal anti-inflammatory drugs (NSAIDs), nitrates, tetracyclines and anticholinergic drug therapy has been significantly associated with an increased incidence of GORD. The past/current use of oral steroids has also been identified as a contributing factor in the development of GORD symptomatology.

There are several comorbidities that have been linked with an increased diagnosis of GORD. These include hiatus hernia, peptic ulcer disease, hypercalcaemia, Zollinger–Ellison syndrome, asthma, irritable bowel syndrome and depression. Physical exercise has been identified as the only protective lifestyle factor in reducing GORD symptomatology. Several studies have indicated that cagA-positive *Helicobacter pylori* infection may have a protective effect against GORD. However, this remains a controversial topic and it should be noted that cagA-positive *H.*

pylori is a known causative factor for both duodenal ulceration and gastric cancer.

Pathophysiology of GORD

Maintenance of the oesophageal mucosal integrity is essential in preventing symptomatic GORD. There are three mechanisms that are thought to be responsible for maintaining this oesophageal mucosal integrity:

- a pre-epithelial defence mechanism consisting of bicarbonate ion, mucus and epithelial growth factors. Submucosal glands are believed to neutralize acid reflux during sleep (note that salivary secretion is decreased at this time)
- resistance to transmucosal ion diffusion due to intercellular junctional complexes and lipid matrices
- the Na^+/H^+ antiporter maintains intracellular pH by inhibition of free-radical production and increasing cell turnover when damage has occurred.

However, once this mucosal barrier is breached, the function of these protective mechanisms is reduced. This creates the environment in which two physiological factors interact to lead to the development of symptomatic GORD. The occurrence of GORD is closely related to transient relaxation of the LOS and increased gastric acid secretion. The failure of the LOS to maintain its role in the antireflux mechanism is the main contributing factor in the pathophysiology of GORD. The resting tone of the LOS pressure is usually between 10 and 25 mmHg. The muscles of the LOS have specific physiological functions and are thought to be mediated via the vagovagal reflex pathway in response to the activation of stretch receptors in the stomach. The LOS is also subject to neuroendocrine influences. Several drugs, hormones and food substances are known to influence the contractility of the LOS. The resting tone of the LOS increases following the administration of gastrin, cholinergic and -adrenergic agents, and decreases with secretin, cholecystokinin and glucagon activity. It has been shown that foods with high fat content, coffee and nicotine from cigarette smoke reduce the LOS pressure.

There are also several factors that may contribute to LOS dysfunction. These include:

- the length of the intra-abdominal segment of the oesophagus
- the diaphragmatic crural mechanism
- the angle of His
- impaired gastric emptying.

The principal physiological components of gastric juice are acid and pepsin. Exposure of the oesophageal mucosa to gastric refluxate, principally gastric acid, is the other major factor in the pathophysiology of GORD. The importance of acid (pH <4) for reflux perception was first demonstrated in a prolonged pH monitoring study by DeMeester *et al.* In this landmark study, the frequency of acid reflux and the mean distal oesophageal acid exposure time were both significantly greater in consecutive patients with typical reflux symptoms (heartburn and/or regurgitation) or dysphagia than in asymptomatic controls. Reflux components other than acid, such as bile from duodenogastro-oesophageal reflux and pepsin, have also been known to contribute to symptom perception in patients with GORD. Pure alkaline reflux is rare and occurs in patients who have undergone gastric resection/bypass or in conditions causing achlorhydia (e.g. pernicious anaemia, somatostatinomas).

The majority of patients with GORD can be categorized into one of three groups with impaired LOS function:

- patients with a hypotensive resting LOS pressure of less than 6 mmHg have free reflux of gastric contents into the lower oesophagus
- patients with a borderline oesophageal sphincter pressure and intermittent increases in intragastric pressure – 'stress reflux'
- patients with an increased frequency and duration of transient LOS relaxations are termed as 'reflex reflux'. This group accounts for 40% of patients with GORD and 65% of patients with reflux oesophagitis.

Pathological changes associated with GORD

The normal oesophageal mucosa consists of three layers: the basal (germinal layer), the prickle cell layer (polygonal cells with numerous bridges) and the superficial or functional layer (flattened cells with pyknotic nuclei). In normal oesophageal mucosa, there are few inflammatory cells and the vascular dermal papillae project from the lamina propria to no more than half the thickness of the epithelium. The histological changes of early mucosal damage include widening of the basal layer so as to constitute more than 15% of the total epithelial thickness. In addition, the dermal papillae extend to more than two-thirds of the way through the epithelium to the luminal surface. These changes indicate an increased rate of epithelial cell turnover. With severe damage, there is accumulation of polymorph inflammatory cells. It has been demonstrated that inflammatory cytokines play an important role in inducing early inflammatory changes in patients with GORD. Specifically, interleukin (IL) 8 and IL-6 produced by leucocytes act together to induce oxidative stress mechanisms which contribute to mucosal inflammation. Studies have also shown that IL-8 gene expression is increased in both patients with GORD and those with non-erosive reflux disease (Figure 22.19).

Pathophysiological and clinical studies have shown that cagA-positive *H. pylori* infection may be protective against the development of symptomatic GORD. However, this protective effect is dependent on the extent of *Helicobacter*-induced corpus gastritis with severe corpus gastritis causing profound reduction of acid secretion.

Symptomatology of GORD

Typical symptoms of GORD include heartburn and regurgitation. When heartburn occurs on more than 2 days of each week, the patient is considered to have severe heartburn and probably GORD. According to the Montreal Consensus group's classification, GORD develops when the reflux of stomach contents causes troublesome symptoms and/or complications (e.g. oesophagitis). The characteristic GORD symptoms included in this statement are retrosternal burning

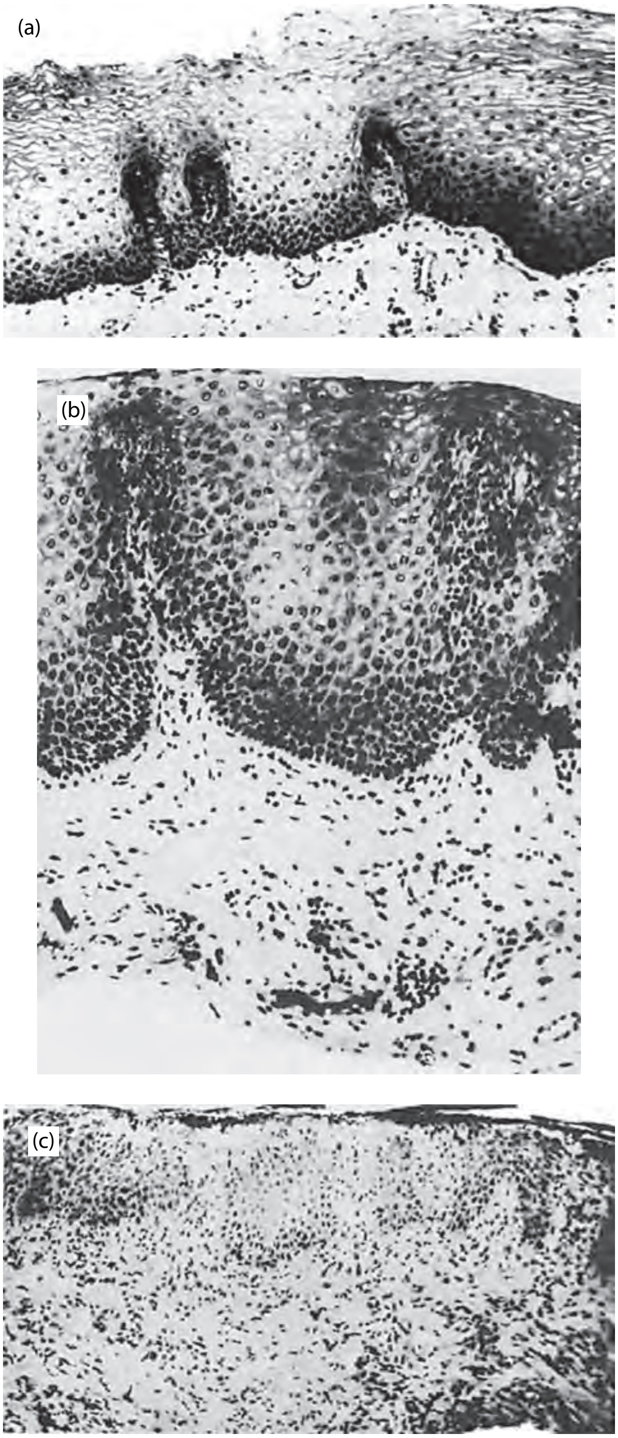


Figure 22.19 (a) Suction biopsy of the lower oesophageal mucosa. Normal histology: the basal layer is less than 15% of the total epithelial thickness. The papillae of the lamina propria extend less than two-thirds of the way to the luminal surface. No polymorphonuclear leucocytes are present in either the lamina propria or epithelium. Mononuclear cells, normal constituents of the lamina propria, are present. (b) Mild to moderate oesophagitis due to acid reflux. The basal layer occupies about 30% of the total epithelial thickness and the papillae extend nearly to the surface. No polymorphonuclear leucocytes are present. (c) Severe gastro-oesophageal reflux with gross erosion and exudate visible endoscopically. The total epithelial thickness is less than normal and is composed entirely of basal cells. The papillae are widened and extend all the way to the luminal surface where superficial necrosis is present. Many polymorphonuclear leucocytes are present in the lamina propria and within the epithelium.

and regurgitation. The term 'troublesome' is specifically included in the definition and is meant to imply that these symptoms impact on the wellbeing of affected individuals (i.e. quality of life).

Heartburn is the most common manifestation of GORD, which may occur in up to 50% of the population. Some patients report severe heartburn, yet on OGD there is little or no evidence of mucosal inflammation. These patients can still have reflux with an abnormal oesophageal mucosal sensitivity. Heartburn is often worsened by a supine position or an increase in intra-abdominal pressure. Heartburn symptoms may also be worsened by fatty meals or intake of alcohol. This symptom can increase in intensity until it is perceived as chest pain.

Regurgitation results from regurgitation of gastric fluid into the throat accompanied by a sour taste in the mouth. Regurgitation is positional and patients will report worse symptoms at night when lying down in bed. Postural regurgitation by meals and activities is associated with a rise in the intra-abdominal pressure (i.e. straining).

Water brash is an uncommon symptom that specifically occurs in patients with GORD. It is due to excessive salivation and the mouth becomes full of fluid.

Other symptoms that can be associated with GORD include dysphagia, odynophagia and chest pain (due to oesophagitis).

Complications of GORD

The complications of GORD are less common since the advent of PPIs. However, they can include reflux oesophagitis, ulcerative oesophagitis, formation of strictures and webs.

Reflux oesophagitis

The development of endoscopically identifiable changes in oesophageal mucosa represents a complication of GORD. A variety of grading systems have been suggested to reflect the severity of the oesophagitis. The most useful grading systems for this purpose include Savary–Miller and the 'Los Angeles classification'. However, a significant proportion of patients with reflux oesophagitis are asymptomatic and monitoring of their disease is important.

Ulcerative oesophagitis

This complication is usually only found in patients with longstanding disease. These patients tend to have more severe reflux and are more symptomatic. The ulcers are beyond the submucosal oesophageal layer and result in perioesophagitis and extensive mural fibrosis. This can eventually lead to stricture formation and oesophageal shortening. Anaemia through chronic blood secondary to ulceration can occur but full thickness perforation is extremely rare.

Oesophageal strictures and webs

Oesophageal webs are the result of submucosal fibrosis in a localized area of the oesophagus, most commonly due to ulcerative oesophagitis (Figure 22.20). They can cause symptoms

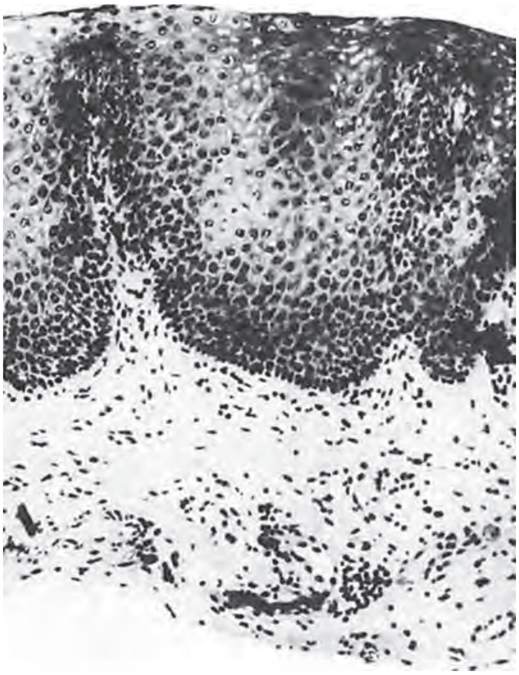


Figure 22.20 Barium swallow showing a Shatski's ring just above the gastro-oesophageal junction causing hold up to the flow of barium in a patient presenting with dysphagia.

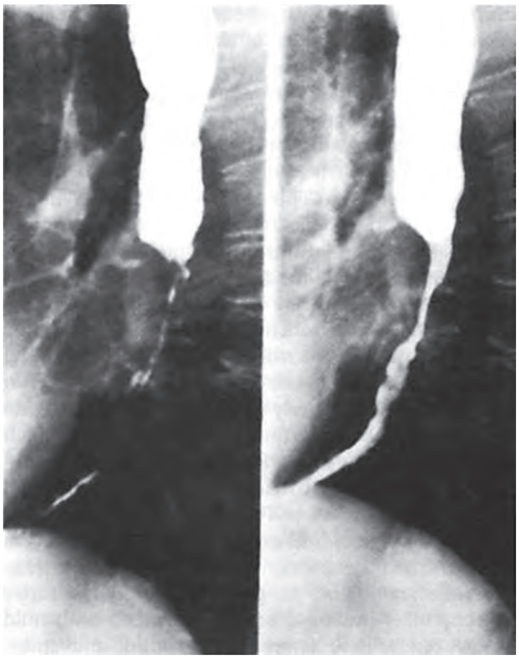


Figure 22.21 Barium swallow showing severe reflux stricture. It exceeds 3.0 cm in length.

of dysphagia and are amenable to endoscopic balloon dilatation. Stricture formation is the result of repeated oesophageal damage with fibrosis replacing the muscular coat of a segment of the oesophagus (Figure 22.21). Oesophageal stricturing occurs in approximately 10% of patients with GORD. The majority of these patients tend to be over 65 years and have a history of longstanding disease.

Treatment of GORD

Treatment of GORD includes a range of options and can be divided into three progressive steps. The first is changes to lifestyle whereby patients are advised to lose weight if overweight, reduce smoking/alcohol consumption and to alter dietary and sleeping patterns. Posture advice includes sleeping propped up to 45°, which can give significant symptomatic relief for those with mild symptoms. The second tier of treatment is the use of acid suppression therapy, using either PPIs or histamine receptor antagonists (H_2-receptor blockers). Acid suppression therapy can be intermittent, but once started on medication the majority of patients with significant GORD remain on long-term treatment. Antireflux surgery is currently considered within the third tier of management. It is usually confined to patients with severe symptoms who have failed medical therapy or whose symptoms are not sufficiently controlled. The treatment choice for GORD is very much dependent on the severity and persistence of symptoms and/or onset of complications. The main aim in the management of GORD is to improve the quality of life in these patients by either complete symptom resolution or maintenance of symptomatic control.

Medical treatment

Medical treatment is indicated when symptoms of GORD are persistent. Antacids (e.g. Gaviscon) can provide symptomatic relief in mild cases. However, neutralization and/or suppression of intragastric acidity remain the mainstay of modern medical therapy. By reducing the acidity, the gastric refluxate produced is less harmful to the oesophageal mucosa. PPIs (e.g. omeprazole, lansoprazole, pantoprazole) are the most commonly prescribed drugs for GORD. They are potent inhibitors of gastric acid secretion, which act specifically by inhibiting the enzyme H^+/K^+-ATPase, the final step in the formation of hydrochloric acid in gastric parietal cells. PPIs will heal 80–90% of patients with endoscopically demonstrable oesophagitis after an 8 week period of high-dose treatment. An alternative to PPIs are H_2-receptor blockers (e.g. cimetidine, ranitidine) and these can be effective in certain patients. H_2-receptor blockers reduce gastric acid secretion by acting as competitive antagonists of histamine at parietal H_2-receptors. However, studies have shown that PPIs are well tolerated, provide the best healing rates for acute oesophagitis and maintain remission better than H_2-receptor blockers.

If a patient has an inadequate PPI response, there are certain factors which should be considered:

- nocturnal acid breakthrough – intragastric pH <4 for more than 1 hour during an overnight period
- *H. pylori* status
- timing of PPI ingestion – decreased bioavailability and variation in drug metabolism
- delayed gastric emptying
- genetic variability in PPI metabolism – rapid metabolism is associated with the 2C19 isoform of cytochrome P450.

A proportion of patients with GORD have impaired peristaltic function and LOS dysfunction. A trial of prokinetic agents (e.g. metoclopramide, domperidone) can provide mild symptomatic relief; however, these drugs have no effect on the healing of oesophagitis.

Surgical treatment

There is evidence to suggest that symptoms of heartburn, reflux and bloating are improved after surgery compared with medical therapy. Overall rates of postoperative complications are low but surgery is not without risk. In the long term, surgery seems to be more cost effective on average than medical therapy. Indications for surgical treatment are as follows:

- GORD refractory to medical treatment
- development of GORD-related complications
- GORD contributing to a decline in respiratory function (e.g. chronic obstructive pulmonary disease, asthma, post lung transplant).

The primary goal of surgery is to improve quality of life for the patient, and therefore careful patient selection is essential. Correlation with oesophageal manometry, pH studies and treatment of a patient's dominant symptoms are the main factors in choosing the appropriate antireflux procedure. The aim of antireflux surgery is to restore an anchor of the oesophagogastric junction below the diaphragm to achieve an adequate length of intra-abdominal oesophagus (minimum 2 cm). In addition to achieving sufficient length, restoration of the competency of the lower oesophageal mechanism is essential for improving GORD symptoms.

A laparoscopic approach is considered the gold- standard technique for antireflux surgery. Studies have confirmed that laparoscopic antireflux surgery is advantageous over the open approach in terms of reduced postoperative pain, earlier mobilization, decreased postoperative morbidity and earlier discharge from hospital. The most commonly performed antireflux operation is the Nissen (360°) fundoplication and it is recommended as the surgical therapy of choice by both the Society of American Gastrointestinal & Endoscopic Surgeons (SAGES) and the European Study Group. There are also several laparoscopic partial fundoplications that can be performed such as Toupet (posterior 270°), Dor (anterior 180–200°) and Watson (anterior 120°). Other less commonly performed operations for GORD include the transthoracic Belsey Mark IV approach, the Hill posterior gastropexy and a Collis gastroplasty (an oesophageal lengthening procedure). A systematic review of five randomized controlled trials on the effect of division of short gastric vessels demonstrated that clinical outcome following laparoscopic Nissen fundoplication appears to be similar regardless of whether the short gastric vessels are divided.

A systematic review of seven randomized controlled trials comparing laparoscopic anterior and posterior fundoplication showed that oesophageal acid exposure time and the prevalence of heartburn are higher after laparoscopic anterior fundoplication than with laparoscopic posterior fundoplication. In the short term this is counterbalanced by less severe dysphagia. However, dysphagia scores become similar in the long term, with a persistent substantial increase in prevalence of heartburn and PPI use after laparoscopic anterior fundoplication. The reoperation rate is twice as high after laparoscopic anterior fundoplication as well, mainly due to reinterventions for recurrent GORD. The prevalence of gas-related symptoms is similar. This systematic review supports the use of laparoscopic posterior fundoplication as the surgical treatment of choice for GORD.

A recent meta-analysis comparing laparoscopic Nissen fundoplication and laparoscopic partial fundoplications demonstrated similar reflux control with both operations. However, Nissen fundoplication was associated with a significantly higher prevalence of postoperative dysphagia and postoperative endoscopic dilatation for dysphagia. A partial fundoplication approach was also associated with a significantly lower reintervention rate, lower dysphagia and gas-bloating symptoms. An individual patient's symptoms and the antireflux experience of the surgeon performing the operation will often dictate the choice of fundoplication.

A Cochrane systematic review of four randomized controlled trials comparing laparoscopic fundoplication with medical therapy found that laparoscopic fundoplication surgery resulted in greater improvements in health-related and GORD-specific quality of life than medical therapy in the short to medium term. There is evidence to suggest that symptoms of heartburn, reflux and bloating are improved after surgery compared with medical therapy, but a small proportion of participants have persistent postoperative dysphagia. Rates of conversion from laparoscopic to open surgery and postoperative complications are relatively low. Surgery also resulted in reductions in acid exposure in the lower oesophagus compared with preoperative values and medical therapy. At present, the evidence suggests that laparoscopic fundoplication surgery is more effective for the treatment of GORD than medical treatment but there are limited data to show whether these benefits are sustained long term. Only one study has assessed patient-reported outcome up to 3 years postoperatively. The costs of surgery are considerably higher than the cost of medical management. Cost data are based on treatment in the first year and therefore the cost and side effects associated with long-term treatment of chronic GORD need to be considered.

Results and complications of antireflux surgery

Symptomatic improvement after antireflux surgery is reported between 65% and 96%. Complete symptomatic resolution is in the range 32–82%. Healing of oesophagitis after antireflux surgery is reported between 63% and 91%. Mortality after antireflux surgery is reported in the range 0.05–0.75%.

Postoperative complications after antireflux surgery include:

- dysphagia
- abdominal bloating
- inability to belch or vomit
- early satiety
- nausea
- diarrhoea.

Recurrent reflux symptoms do occur in patients who have undergone an antireflux operation and it is estimated that about 4% of patients eventually require revision surgery. Studies have shown that laparoscopic revision antireflux surgery is safe and feasible, with good to excellent results reported in 84% of cases. However, laparoscopic revision antireflux surgery is associated with higher risk of conversion, higher morbidity, longer hospital stay and poorer outcomes than primary laparoscopic fundoplication. It is also recommended that laparoscopic revision antireflux surgery be carried out in high-volume centres by experienced surgeons.

Disorders of the diaphragm

Congenital diaphragmatic hernia

The development of the diaphragm is usually complete by the eighth to tenth week of intrauterine life. Complete absence of the diaphragm occurs rarely. Congenital hernias are the result of maldevelopment of the septum transversum. The prevalence of the condition can be up to one in 2100 births and the male to female ratio is 2:1. Approximately 80% of fetuses with congenital diaphragmatic hernias also have polyhydramnios and most cases can now be diagnosed by ultrasonography before the twenty-fifth week of gestation. Attempts at intrauterine diaphragm repair have had limited success. Attempts to temporarily occlude the main bronchus of the hypoplastic lung have also had limited success. In general terms 30% of fetuses with congenital diaphragmatic hernias are stillborn. Fifty per cent of those born alive with a congenital diaphragmatic hernia also have other congenital malformations, most frequently of the nervous system. The association with trisomy 18, 20 and 21 and with Pierre Robin syndrome is also documented. The majority of diaphragmatic hernias are left-sided (75%), some are right-sided (22%) and a few are bilateral defects (3%). The commonest type of diaphragmatic defect is a posterolateral hernia (90%), followed in frequency by eventration of the diaphragm (5%), and the least common defect being a retrosternal hernia (2%).

Posterolateral hernia (through foramen of Bochdalek)

These hernias are posteriorly situated and are due to persistence of the pleuroperitoneal canals, which are the last part of the diaphragm to close. The hernia, which is usually left-sided, presents acutely with respiratory distress in the neonatal period. In adults most of them are asymptomatic. Symptomatic patients present with digestive symptoms due to herniation of the colon, stomach or small bowel.

Parasternal hernia (through foramen of Morgagni)

The diaphragmatic hernia described by Morgagni in 1761 is a rare diaphragmatic anomaly that is nearly always congenital. This hernia is more common on the right and occurs through a triangular anterior defect lateral to the sternum between the sternal and costal attachments of the diaphragm where the superior epigastric artery, veins and lymphatics pass from the chest in to the abdomen. It is usually asymptomatic in the first years of life; it may be discovered by chance on a routine radiograph or cause problems in adult life with episodes of pain and tenderness in the right subcostal region and intermittent obstructive symptoms. Complete intestinal obstruction may supervene. The posteroanterior chest film in these patients shows a rounded gas-containing shadow to the right of the cardiac outline. This shadow is seen to lie behind the sternum on lateral chest films. In doubtful cases, a radiological contrast study is needed to confirm the diagnosis. Surgical treatment is recommended in all cases because of the risk of intestinal obstruction and strangulation. The best approach is through a midline upper abdominal incision. After reduction of the contents in to the abdomen, the sack, which is usually present, is excised and repair is performed by approximating the two diaphragmatic edges with non-absorbable interrupted sutures. Closure with a Marlex or propylene mesh is necessary for large defects. Repair of these hernias can also be performed laparoscopically. This hernia can be associated with cardiac anomalies as in the pentalogy of Cantrell.

Herniation through central tendon

The deficiency in the central tendon may be situated at the apex of the right or left cupola or involve the central part in relation to the pericardium. On the right side, a hernia through the central tendon contains a mushroom-shaped portion of liver parenchyma, which grows through the opening and enlarges on the thoracic surface of the diaphragm. The condition is usually diagnosed accidentally by a routine chest radiograph. It can be easily differentiated from a primary tumour of the diaphragm by ultrasound scanning or three-dimensional imaging. In left-sided hernias the fundus of the stomach usually protrudes as an air-containing cyst on the top of the diaphragm. A central hernia is usually associated with a defect in the pericardium, and small intestine can herniate in to the pericardial cavity. A small defect in the central tendon on the right side does not require any treatment. However, surgical repair of the other two defects is usually recommended because of the risk of mechanical gastric or intestinal complications.

Congenital hiatal hernia

This is usually the sliding type and is associated with gastro-oesophageal reflux. More rarely the hernia is of the paraoesophageal variety. Both can present in infancy and childhood.

Congenital short oesophagus

In the absence of congenital defects, gastro-oesophageal incompetence is often present in the neonate. The condition corrects itself spontaneously during the first few months of life, probably by further development of the intra-abdominal oesophagus. True congenital shortening of the oesophagus in infancy and childhood is very rare. In this condition, the cardia and a large portion of the fundus of the stomach are situated in the mediastinum without any obvious hernial sack or sliding.

Most instances of congenital shortening of the oesophagus are acquired and result from prolonged pathological reflux with fibrosis, ulceration and stricture formation. The fibrosis draws the stomach further in to the chest and, in extreme cases, the oesophageal stricture may be situated at the level of the aortic arch.

Regardless of the aetiology, the most common symptoms are those of spontaneous regurgitation when the infant or child is in the reclining position, recurrent attacks of chest infection or asthma. Aspiration may also result in the development of a lung abscess. The condition requires treatment with an antireflux procedure. Most authorities favour a Collis-type gastroplasty with a partial fundal wrap. Others recommend an intrathoracic Nissen fundoplication.

Eventration of diaphragm

Anomalous congenital development of the diaphragm or its innervation may result in unilateral elevation of the diaphragm. Alternatively, phrenic nerve injury at birth or later, or injury to the diaphragm, may result in the same problem. Differentiation between eventration of the diaphragm and a large congenital hernia, especially of the Bochdalek type, may be difficult or impossible until surgical exploration.

Eventration of the diaphragm has clinical significance only if it is associated with symptoms or when it cannot be differentiated from other serious conditions. The symptoms of eventration, which are identical to those of large congenital diaphragmatic hernias, may occur in the neonatal period and include respiratory distress and tachycardia with impaired cardiac function. In older children, digestive and respiratory symptoms are aggravated by obesity.

In adult patients, the symptoms may be minimal and management is then conservative. Surgical treatment is necessary if symptoms are severe or disabling. Approach is through a left thoracotomy. The procedures which can be used to restore the diaphragm include plication of the diaphragm (Figure 22.22). In severe cases, prosthetic replacement of a much attenuated diaphragm with a synthetic mesh is required.

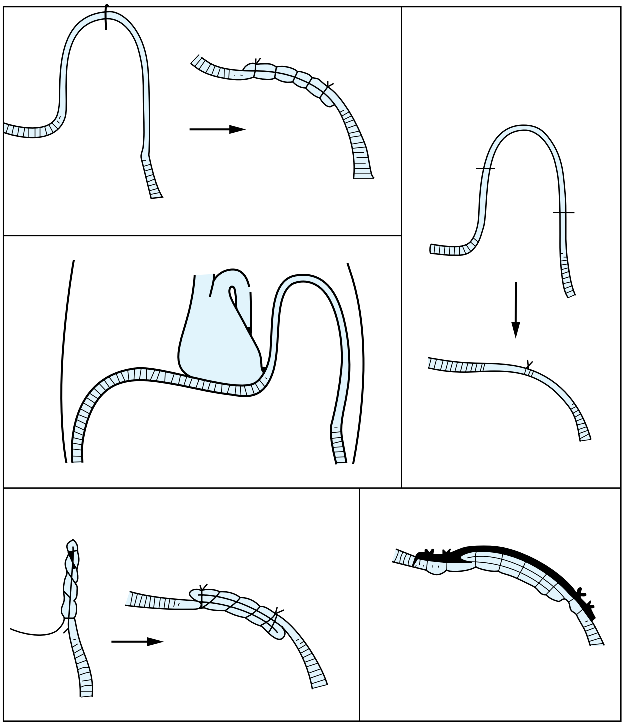


Figure 22.22 Techniques for correction of eventration of the diaphragm.

Traumatic diaphragmatic hernia

Traumatic rupture of the diaphragm may result from penetrating (25%) or blunt (75%) trauma to the abdomen and chest. The tendinous portion, especially on the left side, is the usual side of rupture (68%) as the liver protects the right side of the diaphragm from most injuries except the penetrating type. The rupture is associated with herniation of abdominal contents and may present acutely following the injury or escape detection until several months to years later. The herniation of abdominal viscera may occur acutely at the time of the injury or be delayed until some time later. The symptoms are related to the size of the herniated contents and to the onset of mechanical complications such as intestinal obstruction, strangulation, haemorrhage or progressive cardiorespiratory insufficiency.

The diagnosis is usually established on plain chest and abdominal films when a space-occupying lesion or bowel gas shadow is seen in the chest. If the omentum, spleen or liver is the main herniated structure, the shadow appears solid. The herniated spleen is usually ruptured and accompanied by severe haemorrhage. This may result in total opacification of the left chest. Otherwise, air–fluid levels may be observed indicative of herniation of hollow viscera (colon, small bowel). Passage of a nasogastric tube identifies a herniated stomach above the diaphragm. Contrast radiological studies may be needed to confirm the diagnosis. A significant proportion (40%) of traumatic diaphragmatic hernias do not produce any visible effects on plain chest radiography or three-dimensional imaging using CT, MRI or ultrasonography, making diagnosis elusive in the acute setting. The rupture may be discovered by laparoscopy (gas-free, using abdominal wall lift), thoracoscopy or at laparotomy for associated injuries or subsequently on repeat investigations or as a result of bowel obstruction or strangulation.

In acute rupture, there are often associated injuries, which take precedence over the diaphragmatic injury. However, repair of the acute tear should be performed at the same sitting whenever possible. Elective repair of a traumatic diaphragmatic hernia may be performed through either a left thoracotomy or an upper abdominal approach. In delayed cases, the operation may be difficult due to the presence of adhesions and/or atrophy of the damaged diaphragm. Primary repair is usually possible using interrupted non-absorbable sutures. Otherwise closure with a prosthetic mesh is performed.

Hiatal hernia

The oesophageal hiatus is an elliptical opening in the muscular part of the diaphragm. The crura arise from the anterior surface of the first four lumbar vertebrae on the right and from L2 and L3 vertebrae on the left to insert anteriorly into the transverse ligament of the central tendon of the diaphragm. There is some reported variability in the configuration of the oesophageal hiatus. The diaphragmatic crura are thick, musculotendinous bundles that become more tendinous and more muscular near their vertebral origins. The lowermost portions of the oesophagus and the GOJ are held in place through the hiatus by the phreno-oesophageal membrane. With age, the phreno-oesophageal membrane becomes less definite and more fatty. It is also virtually non-existent in patients with longstanding hiatal hernia. This condition is commonly encountered from the fifth decade onwards in the Western world and there is a strong aetiological association with obesity. Excessive body weight has been found a significant independent risk factor for hiatal hernia. Hiatal hernia, however, is not synonymous with gastro-oesophageal reflux. A hiatus hernia can exist without any symptoms. Further, gastro-oesophageal reflux and reflux oesophagitis can occur in the absence of a hiatus hernia. However, in the presence of hiatus hernia there is probably a higher chance of developing oesophagitis. It is possible that, in the presence of sphincter dysfunction, a hiatus hernia exacerbates reflux disease and its symptoms are worse than in the absence of such a hernia. There are rare instances of post-traumatic herniation of the stomach through the hiatus and these must be differentiated from traumatic rupture of the diaphragm. In the vast majority of cases, however, the development of hiatus hernia is spontaneous. Gallstones and colonic diverticular disease are commonly present in patients with a hiatus hernia (Saint's triad) and difficulty may be encountered in establishing which of the three disorders accounts for the patient's symptoms.

Pathology

Conventionally, three types of hiatal hernias are recognized: type 1, axial, sliding; type 2, paraoesophageal; and type 3, mixed.

Axial hernia

This accounts for the majority (70–80%) of cases. The GOJ and a variable portion of the adjacent stomach slide upward into the mediastinum carrying with them a peritoneal sac. This results in loss of the cardiac angle of His and, commonly, incompetence of the cardio-oesophageal junction. There is no uniform definition of what constitutes a sliding hiatus hernia. Surgeons, anatomists, radiologists and endoscopists all differ slightly in their views and this must be taken into account when evaluating a symptomatic patient. The symptoms and complications of this type of hernia are those which are consequent on gastro-oesophageal reflux and reflux oesophagitis (chronic blood loss, stricture formation, Barrett's epithelium, etc.) (Figure 22.23).

Paraoesophageal hernia

In this type, the fundus of the stomach rotates in front of the oesophagus and herniates through the hiatus into the mediastinum. As the cardio-oesophageal junction remains *in situ* within the abdomen (except in large hernias) cardiac incompetence and reflux are not usually encountered. This type of hernia accounts for 8–10% of cases and is found predominantly in the elderly population. In large hernias the entire stomach and pylorus may be found within the chest inside a large hernial sac, which may also contain the spleen and hepatic flexure of the colon. These large hernias are prone to incarceration and strangulation with infarction and perforation

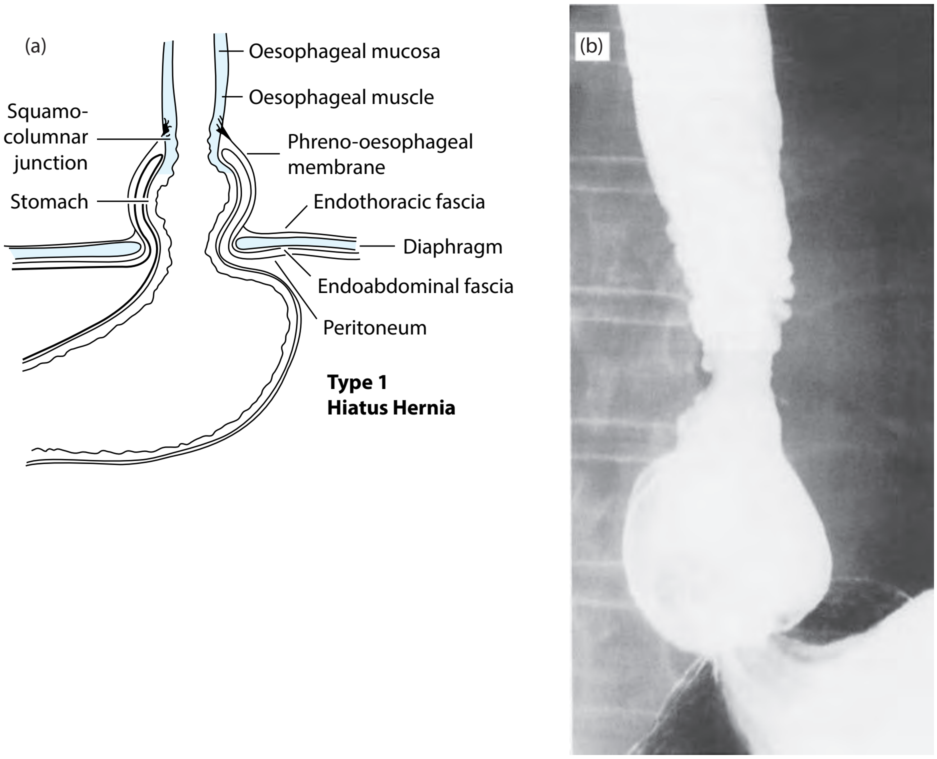


Figure 22.23 Axial (type I) hiatus hernia (a) diagrammatic representation; (b) barium swallow. Note displacement of the cardio-oesophageal junction in the chest.

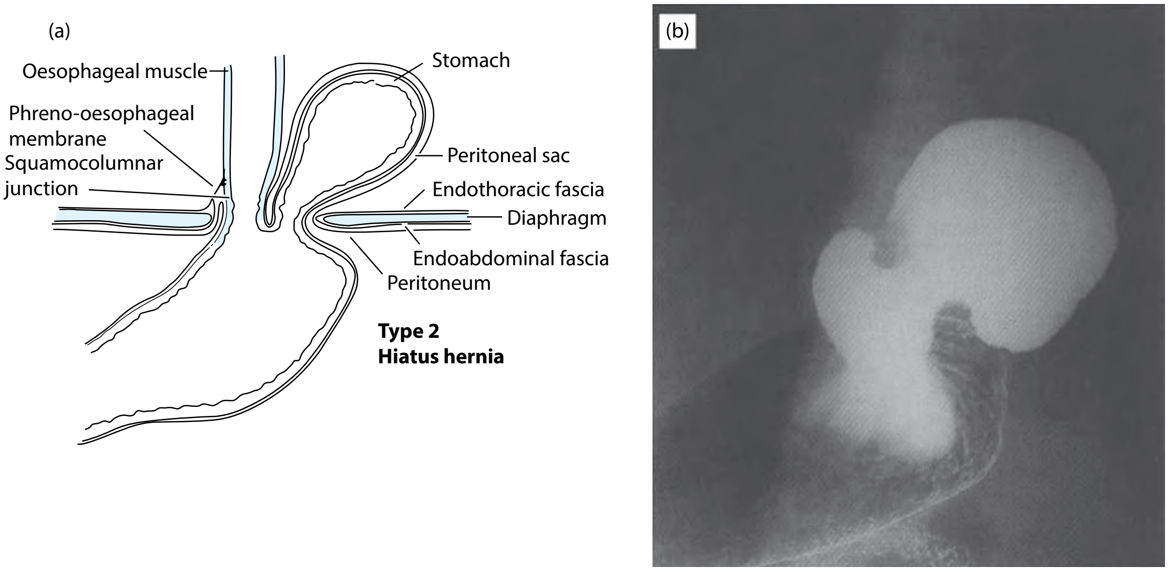


Figure 22.24 Paraoesophageal hernia (type 2). (a) Diagrammatic representation; (b) barium swallow. Note rotation of the fundus as it rolls up in front of the gastro-oesophageal junction.

of the stomach. Large hernias can also progress to complete volvulus, which results in pyloric or duodenal obstruction. When a paraoesophageal hernia bleeds, this is due either to chronic gastric ulceration in the intrathoracic stomach or to an erosive gastritis in a congested and strangulated organ. The majority of uncomplicated paraoesophageal hernias can be easily reduced through the abdomen (Figure 22.24).

Mixed hernia

This resembles a large paraoesophageal hernia but the gastro-oesophageal junction is also herniated above the diaphragm. It has features and complications of both type 1 and 2 hernias. It is found in 10–15% of patients. This type of hernia is generally considered to be a late stage of the paraoesophageal variety (Figure 22.25).

Clinical picture

This depends on the type of hernia and the onset of acute life-threatening complications, which can occur with the paraoesophageal and mixed varieties.

Axial hernia

The condition may be asymptomatic, particularly in elderly patients with limited activity and a sedentary lifestyle. When symptoms occur, they are largely due to gastro-oesophageal reflux and reflux oesophagitis. Chronic blood loss resulting in iron deficiency anaemia is common but active haemorrhage is rare. Some patients may present with dysphagia due to stricture formation without a preceding symptomatic history. Others present with dysphagia secondary to obstruction by diaphragmatic impingement on the herniated stomach.

Paraoesophageal and mixed hernia

The symptoms of paraoesophageal hernias are due to the pressure effects of the herniated stomach, especially when it becomes distended with food or gas. Reflux is rare occurring in only 3% of individuals unless the hernia is or becomes mixed. Common symptoms include pain, dyspnoea, feeling of distension and tiredness, which are precipitated by meals, bending and stooping. The pain is sharp, situated beneath the lower sternum and radiates to the back. It is often accompanied by a bloated sensation, anxiety, palpitation and dyspnoea. The attacks may simulate angina pectoris very closely, and even cardiac arrhythmias may be present during an episode. The pain, however, is often relieved by belching or vomiting. Dysphagia is found in 20% of patients with paraoesophageal hernia.

Acute presentation

Approximately 20% of patients with large paraoesophageal/mixed hernias may present acutely with severe upper gastrointestinal haemorrhage or strangulation/infarction/perforation of the intrathoracic stomach. In the latter instance the patient develops severe retrosternal pain and shock, which are often mistaken for myocardial infarction. A chest radiograph shows a large gastric gas–fluid shadow overlying the heart. With gastric infarction and perforation, mediastinal widening and emphysema, left basal collapse and pleural effusion may be outlined by this investigation. Gastric infarction and perforation carry a high mortality rate from septic mediastinitis and bacteraemia.

Management

Clinical assessment and appropriate investigations must establish that the symptoms are due to the hiatal hernia. In elderly patients and in individuals with comorbid disease, case selection for surgery requires clinical judgement based on the severity of the symptoms and cardiorespiratory reserve. Middle-aged patients with significant coronary artery disease may require myocardial revascularization before surgical treatment of the hiatal hernia.

Type 1, axial hernias are treated by reduction with an antireflux procedure. This is best achieved by a laparoscopic procedure in the majority of patients. The majority of uncomplicated paraoesophageal hernias can also be approached similarly and are easily reducible via this approach. Following reduction of

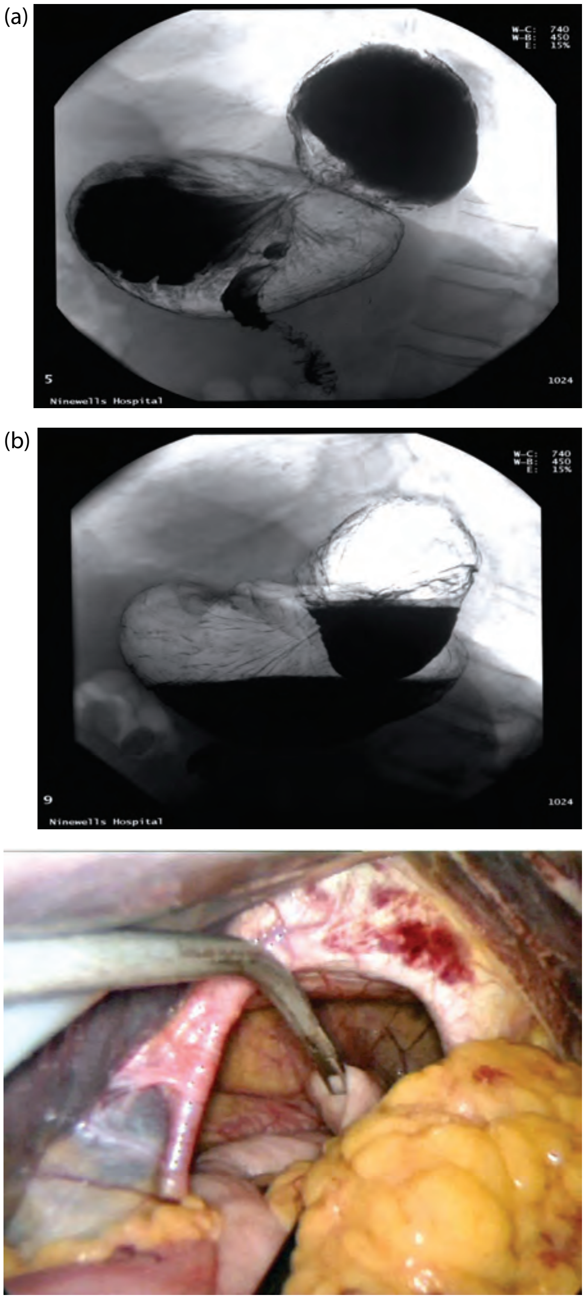


Figure 22.25 Mixed (type 3) hiatal hernia. (a,b) Barium swallow showing large totally intrathoracic herniated stomach with organoaxial volvulus; (c) laparoscopic view of the same patient. The herniation involved the whole stomach, small intestine and spleen. These hernias are particularly prone to strangulation.

the hernia, a small or moderate-sized hiatus is repaired with interrupted non-absorbable sutures and the GOJ fixed beneath the diaphragm after restoring the oesophagogastric angle. Some surgeons advocate a Nissen fundoplication in addition to reduction and crural repair of these hernias. If the hiatal defect is large, a synthetic mesh can be fashioned around the oesophagus and sutured to the edge of the large defect (Figure 22.26). In addition to the above, some surgeons advocate a gastropexy in the form of a tube gastrostomy or otherwise to prevent recurrence. These patients can be managed by the laparoscopic approach by experienced surgeons. They are, however, challenging cases in time and effort.

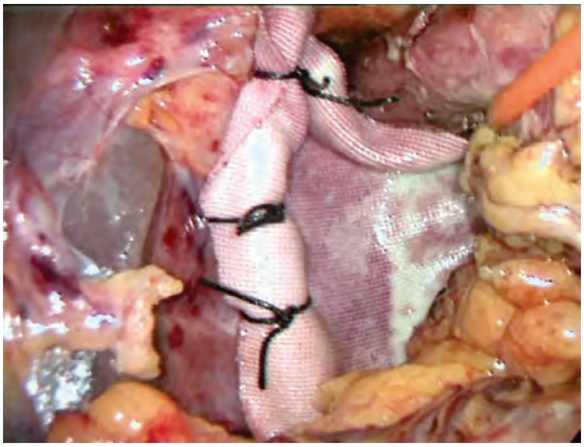


Figure 22.26 Mesh repair of a crural defect in a giant hiatus hernia. The operation is completed by total fundoplication and an anterior gastropexy. Even so recurrence occurs in up to 30% but most are asymptomatic.

Strangulated/infarcted paraoesophageal and mixed hernias require an emergency thoracotomy. If the stomach is viable it is unrotated and reduced into the abdomen and crural repair is performed. Resection of the infarcted stomach with mediastinal and pleural lavage is necessary for those patients presenting with this serious complication.

Oesophageal diverticula

Oesophageal diverticula are an uncommon pathology that can occur at any level of the oesophagus. They are predominantly an acquired pathology caused by pulsion or traction.

Oesophageal diverticula are defined anatomically or by the mechanism which produces them:

- Anatomically
  - upper oesophagus (pharyngo-oesophageal junction)
  - mid-oesophagus (parabronchial)
  - lower oesophagus (epiphrenic).
- Mechanism of formation
  - Traction – chronic inflammation of the mediastinum, particularly mediastinal lymph nodes, involves the mid-oesophagus wall. Fibrosis eventually occurs between the two structures, resulting in the oesophageal wall (mucosa and muscle) being pulled outwards to form a pouch. They are usually small, wide necked and clinically non-significant, producing few (if any) symptoms.
  - Pulsion – an increased intraluminal pressure forces the oesophageal mucosa to herniate, resulting in diverticula formation. Herniation occurs in regions of the oesophagus which are weakened or have a low resistance. Killian's triangle, based in the posterior pharynx and bounded by the cricopharyngeal and thyropharyngeal muscles, is an area of natural weakness and frequently the site of pharyngo-oesophageal (Zenker's) diverticula. Alternatively, oesophageal dysmotility syndromes such as achalasia can alter oesophageal peristalsis, creating uneven pressures and mucosal tears, where again diverticula can develop.

The clinically important diverticula can be divided as follows:

- pharyngo-oesophageal – Zenker's diverticulum (65%)
- mid-thoracic (15%)
- lower oesophagus – epiphrenic diverticulum (20%).

In this section we will primarily discuss the pharyngo-oesophageal and epiphrenic diverticula together with their investigation and treatment.

Pharyngo-oesophageal (Zenker's) diverticula

Zenker's diverticulum is a pulsion-type diverticulum that forms in the midline of the neck, usually in the naturally weak Killian's triangle (see above). The UOS, predominantly formed by the cricopharyngeus muscle, relaxes on swallowing to allow food into the oesophagus and also contracts to prevent backflow into the pharynx. Although there is considerable controversy as to the exact pathology, dysfunction of the sphincter is widely believed to generate excessive pressures which coupled with the weakness in Killian's triangle results in forming Zenker's diverticulum.

The outpouching that forms is mainly composed of the mucosa and submucosa layers and an incomplete muscular coat. Over time, the diverticulum progressively enlarges. However, as the spine prevents indefinite posterior diverticular extension, it instead displaces the oesophagus and comes to lie on the side of the neck (usually on the left) in line with the pharynx (Figure 22.27).

As a result of this anatomical repositioning, food can occupy the diverticulum, where it is retained and be subsequently associated with multiple symptoms and complications.

Clinical features

Box 22.4 shows some of the clinical signs and complications that can be associated with pharyngo-oesophageal diverticula. As can be appreciated, the diffuse, non-specific and highly variable nature of the symptoms makes a clinical diagnosis near impossible, except for in the most advanced cases when the pronounced swelling of the pouch may be palpable and associated with gurgling sounds. As a result, definitive diagnosis relies heavily on imaging of the oropharynx and oesophagus.

Diagnosis

Barium swallow

Barium meal study with fluoroscopy is regarded as the first-line investigation for suspected cases. This should show the definitive swelling of the condition in the pharyngeal area and exclude any advance oesophageal motility disorder or hiatus hernia whose symptoms can overlap with those of oesophageal diverticula (Figure 22.28).

Further tests

Endoscopy, manometry and pH monitoring are occasionally used as adjunct tests in suspected cases. Endoscopy in particular must be undertaken with great care as there is a documented risk of iatrogenic perforation of the diverticulum. It does, however, allow the entire length of the upper gastrointestinal tract to be visualized and exclude gross organic pathology, reassuring both the patient and the surgeon. Manometry helps to exclude an oesophageal motility disorder, which may or may not be present, whereas pH monitoring helps to exclude GORD.

Treatment

Definitive treatment of Zenker's diverticulum is always interventional but dependent on the severity of symptoms, the patient's willingness to undergo surgery and indeed the degree of comorbidity. However, in the modern era, all but extremely frail patients should be suitable for endoscopic treatment under sedation. Patients can be treated equally well with open or minimal access approaches. However, on balance, it seems that high-risk patients with medium-sized diverticula are likely to benefit more from endoscopic/fibreoptic diverticulostomy whereas those with small or giant diverticula are probably best served by open surgery.

Open approach

The open approach is through an incision at the anterior margin of the sternomastoid muscle to the anterior end of the clavicle on the side at which the diverticulum is most prominent (usually the left).

As the UOS is considered defective, a myotomy is performed of the cricopharyngeus muscle, which by itself may be sufficient to relieve symptoms in cases of small diverticula. If the diverticulum is large, then the myotomy is combined with a diverticulectomy or diverticulopexy.

Endoscopic stapling diverticulostomy

Under general anaesthesia, a diverticuloscope with two blades is placed through the pharynx into both the oesophageal lumen

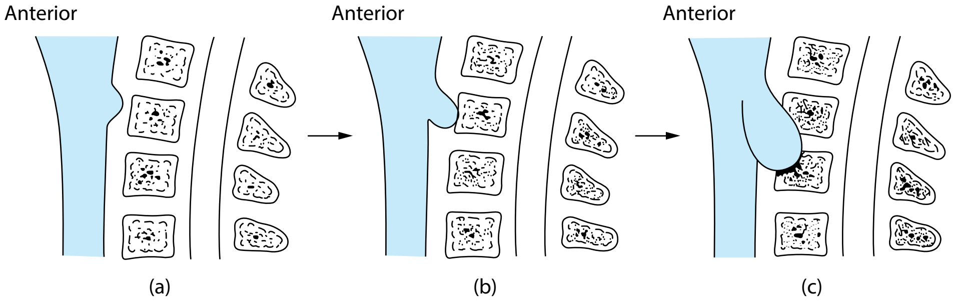


Figure 22.27 Diagrammatic representation of the development and progression of a pharyngo-oesophageal diverticulum. The cricopharyngeal dysfunction and resultant high intraluminal pressure cause a mucosal herniation between the oblique fibres of the inferior constrictor and the transverse fibres of the cricopharyngeus (a). Enlargement of the pouch is limited posteriorly by the spine (b). It therefore becomes dependent and bulges lateral to the oesophagus usually on the left side (c). As it becomes dependent, food tends to enter the sac rather than the oesophageal lumen, which may be compressed by the pouch.

and the diverticulum. A smaller, 5 mm endoscope is passed through to visualize the septum dividing the oesophagus and diverticulum. Prior to stapling, the length of the diverticulum is measured with a graduated rod. Stapling is used to bring together the adjoining walls of the oesophagus and pouch that form the septum, which is then transected forming a diverticulostomy. Cautery is also occasionally used in place of stapling to achieve the same result. The procedure is fast, taking around 20 minutes to perform in skilled hands but nonetheless involves a general anaesthetic.

BOX 22.4 Clinical features and complications of Zenker's diverticulum

- M>F (3:1), middle aged/elderly more affected
- Dysphagia – due to food being stuck within the pharynx and oesophageal compression
- Cough
- Throat irritation
- Food regurgitation of non-acidic material (food is from the pouch)
- Food aspiration – this can lead to pulmonary complications such as pneumonia
- Neck swelling (if the pouch is sufficiently large) with gurgling
- Halitosis (food stasis and fermentation)
- Anorexia
- Higher risk of perforation on endoscopy
- Bleeding
- Carcinoma (0.3%)

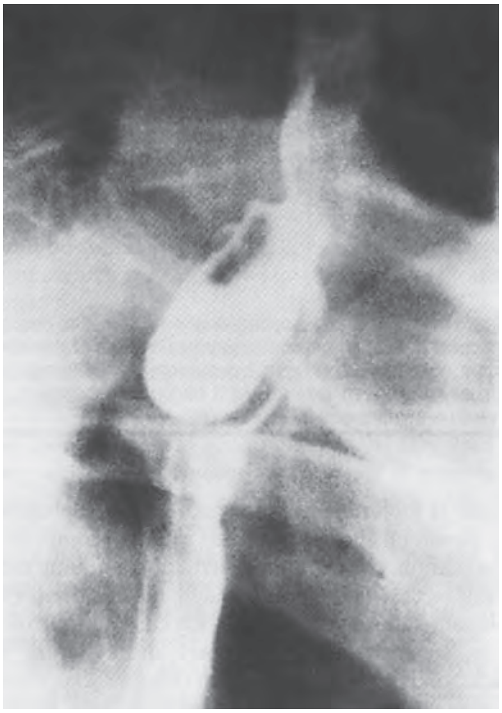


Figure 22.28 Barium swallow demonstrating a large pharyngeal pouch accompanied by an aspiration lung abscess.

Fibreoptic endoscopy

This procedure is performed under sedation and is an alternative to endoscopic stapling under general anaesthesia, making it particularly useful for treating the frail and elderly populations. A fibreoptic endoscope is used to visualize the dividing septum, which is then obliterated using cautery, argon plasma coagulator or CO_2 laser and ends in formation of a diverticulostomy. Diverticula <2 cm in size can be managed in a single sitting with this technique but larger pouches may require repeated procedures.

Complications

Endoscopic treatment is faster and less invasive than open surgical therapy with fewer reported complications. However, whereas the results of surgical myotomy are highly favourable with a good outcome in almost all patients, with endoscopic therapy good outcomes are more variable between 75% and 100%. Additionally, all invasive modalities have associated complications (Box 22.5).

Mid-thoracic diverticula

Mid-thoracic diverticula are the least common of the oesophageal pouches and can arise as a result of pulsion (secondary to an oesophageal motility disorder causing increased intraluminal pressure and wall weakness) or traction (from fibrotic adhesions between the oesophageal wall and mediastinal structures such as inflamed lymph nodes). The diverticula tend to possess a narrow neck and a limited muscular coat. Asymptomatic or minimally symptomatic patients are usually managed conservatively, whereas those with extensive pouches are surgically treated by excision and layered oesophageal closure through a right-sided thoracotomy.

Epiphrenic diverticula

Although the second most common of the oesophageal diverticula, epiphrenic diverticula are nearly five times less common that Zenker's pouches, which are themselves infrequent pathologies. The great proportion of epiphrenic diverticula are acquired as a result of raised intraluminal oesophageal pressure secondary to a motility disorder such as DOS or achalasia. Occasionally, they may arise in conjunction with a hiatus hernia or gastro-oesophageal reflux. The symptoms are largely due to the underlying disorder, although ulceration and bleeding within the diverticulum has been documented as a rare cause of haematemesis (Box 22.6).

BOX 22.5 Complications of surgical therapy for Zenker's diverticulum

- Haematoma
- Recurrent nerve palsy
- Suture line leakage
- Cardiovascular complications (especially in elderly/frail patients)

BOX 22.6 Clinical features of epiphrenic diverticula. As with other oesophageal diverticula and oesophageal conditions in general, many of the symptoms are shared making a purely clinical diagnosis very difficult

- Dysphagia
- Cough
- Food regurgitation (occurs in 50% of patients who seek medical attention)
- Retrosternal chest pain
- Heartburn
- Halitosis (food stasis and fermentation)
- Anorexia
- Ulceration and bleeding
- Carcinoma
- Spontaneous perforation (similar to Boerhaave syndrome)

Diagnosis

Barium swallow, as in Zenker's diverticula, will show the clear outline of the pouch in the lower part of the oesophagus. Given its association with motility disorders, manometry is an important part of the diagnostic work-up, which can show tertiary contractions and further evidence of dysmotility such as aperistalsis or failure of relaxation of the LOS. Standard manometry fails to show dysmotility in around 25% of cases, but prolonged manometry may pick up more subtle dysmotility.

Treatment

Treatment is dependent on patient severity and the underlying condition that has produced the diverticulum. Generally, asymptomatic or minimally symptomatic individuals are managed conservatively and may be imaged infrequently to ensure that the diverticulum is not growing to a disproportionate size or beginning to cause symptoms.

Additionally, if an underlying motility disorder or reflux disease has been identified, it is important to treat it as far as possible, thereby removing the causative agent of the epiphrenic diverticulum's formation and preventing further enlargement. In cases where the diverticulum is large or symptomatic, i.e. compressing the oesophagus, or is unable to drain or inflamed, then surgical intervention is the preferred approach with the primary aim being to excise the diverticulum and, if applicable, correcting any underlying disorder that has contributed to pouch formation.

Traditionally, surgical treatment involves:

- diverticulectomy
- oesophageal myotomy – if an underlying motility disorder such as achalasia has been identified
- antireflux procedure – e.g. Door partial fundoplication to prevent gastro-oesophageal reflux post myotomy.

Thoracotomy/thoracoscopy

Prior to the laparoscopic era, a right-sided thoracotomy was the preferred surgical approach, as this allowed sufficient access to the offending pouch to be excised and to the oesophagus to perform a myotomy.

This preference was carried into the laparoscopic era, where a four-port right-sided thoracoscopy was performed. The epiphrenic diverticulum would be identified intraoperatively by endoscopy and then isolated by blunt dissection. The diverticulectomy would be performed by staple application at its base followed by a layered closure of the oesophagus. The endoscope could then be used to test the efficacy of the closure and identify any missed perforations.

Although this approach had a symptom success rate of 85%, there was a risk, as with all thoracotomy/thoracoscopy, of staple line leak and mortality. Furthermore, the complex nature of the anaesthetic, positioning of the patient and inaccessibility to the stomach to perform a fundoplication has resulted in the technique being largely abandoned for the laparoscopic approach.

Laparoscopy

Abdominal laparoscopy is also performed using a four-port technique to access the hiatal structures and the diverticulum within the mediastinum, staying close to the oesophageal wall. Once the diverticulum is isolated, its neck is sutured and a linear endoscopic stapler applied to transect the pouch. Prior to the transection, the oesophagus can be inflated endoscopically or a bougie placed in order to prevent luminal narrowing during transection.

If oesophageal dysmotility is identified preoperatively, a myotomy is then performed on the lower oesophagus extending for 2–2.5 cm on to the gastric cardia. As gastro-oesophageal reflux is the main complication of a myotomy, a Dor anterior fundoplication is additionally fashioned. Overall results of this procedure are excellent with much greater gastric access, less morbidity and a faster recovery than the thoracoscopic approach.

Oesophageal injury

Oesophageal perforation

Oesophageal perforation is traditionally regarded as a surgical emergency with very high associated morbidity and mortality rates of 10–40% reported. Treatment of oesophageal perforation remains an area of interest and debate for oesophageal surgeons. There is now consensus that patients with oesophageal perforations are best treated with early diagnosis and referral to specialist tertiary centres.

Aetiology

- Instrumental
- Intraoperative iatrogenic injury
- Trauma
- Boerhaave syndrome
- Foreign body

The commonest cause of oesophageal injury is instrumental following upper gastrointestinal endoscopy. The incidence of oesophageal perforation following insertion of a rigid endoscope is 0.5%. Rigid endoscopy is rarely used in current gastrointestinal clinical practice but it is still used by otolaryngology surgeons. Oesophageal perforation may occur following therapeutic endoscopy for dilatation or intubation of strictures. The incidence of perforation is 1–5% after pneumatic dilatation for achalasia using 4 cm balloons. Endoscopic injuries also occur during EUS and endoscopic retrograde cholangiopancreatography (ERCP). Oesophageal intubation with a transoesophageal echo probe during cardiac ablation is another cause of oesophageal injury. The commonest site of injury during therapeutic endoscopy is the location of intervention, i.e. stricture, whereas cervical oesophagus is the commonest site for injury during EUS, ERCP and echo cardiography. Oesophageal diverticulum is an important predisposing factor of cervical iatrogenic endoscopic injury.

Intraoperative perforation occurs due to oesophageal damage sustained during paraoesophageal surgery, e.g. fundoplication, repair of hiatus hernia, thoracic aortic surgery and bariatric surgery. It is a technical error during surgery. Redo surgery, poor surgical exposure and inadequate surgical expertise are important factors that underline such technical errors.

Trauma of the oesophagus may occur due to penetrating gun shot or stab wounds. The cervical oesophagus is the segment most commonly involved. It is usually associated with injuries to adjacent structures.

Boerhaave syndrome is a serious condition of forceful vomiting followed by severe chest pain due to a complete tear of the lower thoracic oesophagus, just above the cardia. It was first described by Herman Boerhaave in a Dutch admiral. This aristocratic gentleman succumbed in this way following a bout of overindulgence of food and drink. However, only a minority of complete spontaneous perforations of the lower thoracic oesophagus fit the classical description of Boerhaave. The condition is very uncommon and occurs usually between the ages of 40 and 60 years with a male to female ratio of 2:1. There is frequently a long history of indigestion and chronic gastrointestinal disease such as duodenal ulcer, reflux oesophagitis and hiatal hernia. Apart from overeating, other predisposing factors include neurological disorders, tumours and gastrointestinal obstruction.

Ingestion of a foreign body such as a fish bone is a rare cause of oesophageal perforation. It is usually difficult to recognize and, on many occasions, it can only be diagnosed retrospectively on taking a detailed history or after surgery.

Clinical picture

The clinical presentation depends on the cause and onset of perforation, site of perforation and the extent of perforation and subsequently the degree of contamination and its systemic effect.

The acute manifestations of an oesophageal perforation are severe pain, dyspnoea, tachycardia and fever. The site of the pain and its radiation vary with the oesophageal segment involved. Haematemesis may also be found in cervical perforations. Subcutaneous emphysema is observed in 60% of cervical and 30% of mid-oesophageal injuries. In thoracic injuries, respiratory distress is common due to effusion and the systemic effect of sepsis due to mediastinitis. Upper abdominal tenderness indicates perforation of the abdominal segment of the oesophagus or gastro-oesophageal junction with oesophageal contents tracking into the abdomen. On the other hand, pain may be the only symptom in minor injuries and may not be associated with systemic signs of sepsis. The onset of symptoms in relation to the underlying cause is usually noticeable specially following endoscopic interventions or paraoesophageal surgery. In non-iatrogenic perforation, the clinical picture may be misleading as pain, sepsis and respiratory compromise may simulate acute chest conditions.

In late presentation, the clinical picture is that of established sepsis with its systemic cardiovascular and respiratory effects. A detailed history will elicit the early manifestations that are misdiagnosed with other conditions. In this group of patients, oesophageal perforation is usually diagnosed during investigations for sepsis. It is essential that clinical history should include evaluation of the onset of perforation, comorbidity, quality of life and the degree of patient dependency in old patients. In case of oesophageal strictures, determination of the underlying pathology is important. Those factors are crucial in deciding management strategy.

The diagnosis is usually confirmed by imaging. Plain radiographs are frequently diagnostic but cannot accurately localize the site of perforation. The radiological features include the presence of surgical emphysema in the mediastinum or neck, widening of the mediastinum or mediastinal air at the cardiac border (Figure 22.29). Free air under the diaphragm may be detected in patients with injuries to the abdominal oesophagus. CT scan of neck, chest and upper abdomen with intravenous and oral contrast is always required. It determines the level of perforation and the presence of collection in the mediastinum or communication with the pleural cavity. In cases of minor perforations, only mediastinal air is seen with no evidence of extravasation of contrast. A water-soluble contrast swallow is helpful to localize the perforation and the side of perforation, which is important in determining the surgical approach if needed (Figure 22.30). Endoscopy is required if the surgical option has been decided because endoscopy is the most accurate investigation to determine the extent of mucosal injury, which is often larger than the size of muscle tear.

Investigations should include blood count and biochemistry and cardiac and respiratory assessment as part of the management plan. Frequent blood gases and lactate levels are important. In cases requiring surgical interventions and those with cardiorespiratory compromise, early involvement of an anaesthetist and intensive care team is important.

Management

The management of oesophageal perforations consists of (1) optimum resuscitation of patients, (2) establishing adequate nutrition, preferably via the enteral route, and (3) control of sepsis.

Adequate resuscitation is a key component in the management of oesophageal perforations whether surgical or conservative approaches are adopted. Patients with cardiorespiratory

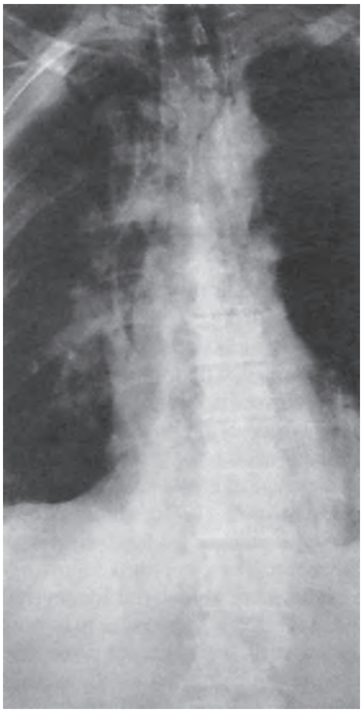


Figure 22.29 Plain radiograph of the chest showing surgical emphysema in the mediastinum following oesophageal perforation.

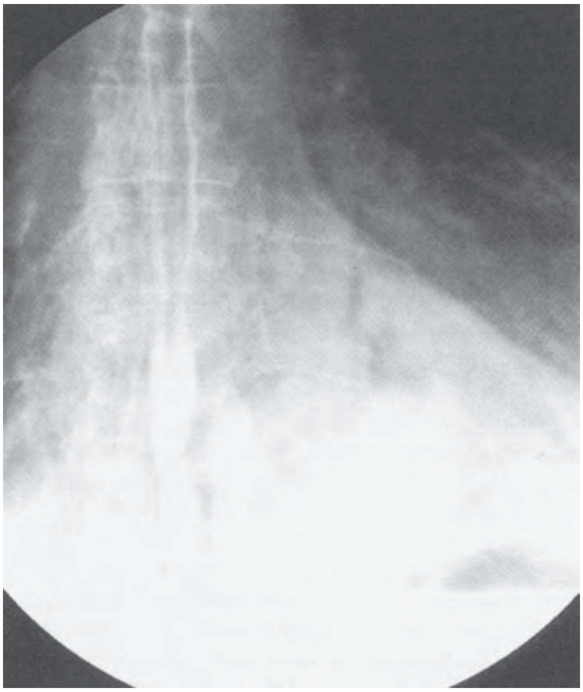


Figure 22.30 Water-soluble contrast swallow confirming and outlining the site of perforation in the lower oesophagus.

compromise or severe sepsis should be treated in high-dependency units or in the intensive care unit. Monitoring the response to management and the early detection of any deterioration are essential.

Nutrition should be established at an early stage. In patients requiring surgery, insertion of feeding jejunostomy provides an enteral route that is easy to manage and allows the patients to go home on supplementary feeding if required. In patients with minor perforations who are expected to recover within a few weeks, fine-bore nasojejunal or nasogastric tubes (in cases with no reflux) provide a short-term enteral route. Total parenteral nutrition may also be an option in some cases until the enteral route is established.

Control of sepsis needs antimicrobial therapy with broad-spectrum antibiotics (which includes anaerobic cover) and antifungal medications. This should start as soon as a diagnosis is made. Drainage of collections is essential. Drains into the pleural cavity or radiologically guided drains into the chest/abdominal abscess cavity are important for sepsis control. The aim of surgical interventions is to repair the perforation in order to control sepsis. Patients need to remain nil by mouth with drainage of the stomach via nasogastric tube or gastrostomy until healing of the perforation is demonstrated.

Conservative approach

The indications for a conservative approach are:

1. minor perforations with either no leakage of contrast (only mediastinal air demonstrated on CT) or a small contained leak
2. controlled sepsis with no signs of systemic cardiorespiratory compromise
3. no underlying pathology or distal obstruction
4. in late presentations where surgical repair would be associated with increased morbidity
5. in high cervical injuries below the cricopharyngeus where surgical access would be limited for adequate repair.

The conservative approach is sometimes forced to be adopted because of patient frailty and comorbidity that make surgical options likely to be unsuccessful.

The principles of conservative management are detailed above and consist of optimum resuscitation, adequate nutrition and control of sepsis. Adequate drainage of sepsis needs monitoring of patient condition and imaging the progress of the size of residual collections related to the leak. New collections or inadequate drainage need insertion of new drains. Insertion of a double-lumen drain provides a means for continuous irrigation as a closed irrigation system for contained cavities.

Endoscopic therapy of oesophageal perforations

The endoscopic treatment of oesophageal perforations can be used as part of the conservative approach. Application of metal clips or sealing glue may be used. These methods work when the edges are not fibrotic or very friable and hence they are more suited for early small perforations.

Oesophageal intubation with covered self-expanding metallic stents is gaining popularity for sealing oesophageal perforations in some centres. It only works in perforations of the thoracic oesophagus. Perforations within 2 cm of the cricopharyngeus or in dilated oesophagus are not suitable for stent placement. Oesophageal stents are usually used for benign perforations. In perforation of malignant strictures, stent placement is considered either as a salvage procedure to bridge the time before definitive surgery or as a palliative procedure because tumour perforation is considered as a local spread. Oesophageal stents can be used for early- and late-diagnosed perforations and for small or large tears. Temporary drainage of gastric contents can be achieved by a percutaneous endoscopic gastrostomy, which can be performed at the same sitting and through this aperture a jejunal feeding tube can be introduced and directed through the pylorus for nutritional support.

For selection of stents, the non-flared portion of the stent should be more than 21 mm if no stenosis is associated with the perforation. The length of the stent should be 12–15 cm to adequately cover the perforation and prevent reflux into the perforation.

Surgical treatment

The overall aim is to seal the perforation, clear gut contents extravasated outside the lumen, drain the site of perforation, drain the stomach to minimize the reflux into the repair site and maintain an enteral route for nutrition usually via a feeding jejunostomy.

Factors that influence surgical treatment

- Onset of presentation
- Level and site of perforation
- Underlying pathology – cause of perforation
- The degree of established local sepsis
- Size of perforation – quality of adjacent oesophageal tissue.

The best results are obtained with early perforations as the mortality increases significantly when perforation is treated after 48 hours of onset. Early presentation usually allows direct repair because of the adequate quality of adjacent oesophageal tissue and the ability to control local sepsis.

Surgical approach

The surgical approach depends on the level (cervical, thoracic, lower oesophageal) and the site (right or left) of oesophageal perforation. *Cervical injuries* are approached through an incision along the anterior border of the left sternomastoid muscle whereas *thoracic perforations* are approached through a right or left thoracotomy (depending on the exact level and site of perforation). *Lower thoracic* and *abdominal oesophageal perforations* are approached via the diaphragmatic hiatus through a midline laparotomy or via a left thoracoabdominal incision in cases of long oesophageal tears.

Oesophageal repair

There are several options on how to deal with the perforated oesophagus. An upper gastrointestinal endoscopy is essential to accurately determine the size of mucosal tear, which usually exceeds the size of muscle tear.

1. *Direct repair* entails refreshing the edges of the tear to realize good quality tissue that enables direct suturing usually with 3/0 PDS. Some surgeons recommend buttressing the thoracic oesophagus with pleural or intercostal flaps for high perforations and a diaphragm flap, pericardial patch or omental patch for low repair. Repair of the abdominal oesophagus may be reinforced by an anterior fundal patch or a Nissen fundoplication in posterior injuries. The risk of narrowing the lumen can be minimized when repairing the oesophagus while the endoscope is left *in situ*.
2. *Suturing the perforation over a T-tube (silicon, size 22–24 French)* can be used in cases of late presentation with poor quality of adjacent oesophageal tissue and established local sepsis that cannot be controlled with local drainage. This option is preferred over resection in patients with poor comorbidity who cannot withstand the impact of oesophagectomy.
3. *Oesophagectomy* can be carried out in cases of perforations with underlying pathology that impair healing, malignant perforations, and in late presentations with established local sepsis. Patient comorbidity and clinical status is vital in deciding oesophageal resection. Joining the oesophagus with the stomach conduit to restore gut continuity should be carried out in the neck to avoid having the anastomosis in the presence of local sepsis. Primary reconstruction is preferred.
4. *Oesophageal diversion with neck oesophagostomy* will be performed in cases of uncontrolled sepsis with local measures of drainage. This may be accompanied with oesophagectomy depending on the clinical status of the patient. Reconstruction of gut continuity with colonic interposition can be performed at a later stage after the patient recovers from sepsis and improves his/her nutritional status (Figure 22.31).

Oesophageal repair is the first step towards control of sepsis. Good lavage of the surgical field is essential. Pleural and mediastinal drains should be inserted.

Mallory–Weiss syndrome

This syndrome consists of painless haematemesis after vomiting and straining induced by, usually, excess alcohol intake. There is a high incidence of associated gastro-oesophageal disease. The Mallory–Weiss syndrome is common and accounts for 5–10% of patients undergoing endoscopy for haematemesis.

The lesion consists of a longitudinal mucosal tear involving the mucosa alone or the mucosa and submucosa on the gastric side of the oesophagogastric junction. The tear, which may be single or multiple, is located on the lesser curve side in the majority of cases (85%). Associated lesions are found in 75% of patients and include hiatal hernia, oesophagitis, oesophageal varices and duodenitis/peptic ulceration. Although the bleeding stops spontaneously in the majority of patients, it may be severe and recurrent.

The condition is more often found in males with a history of alcoholism but not invariably. The diagnosis is confirmed by upper gastrointestinal endoscopy. The treatment is conservative with gastric acid suppression. Endoscopic therapy is reserved for patients with actively bleeding tears at the time of endoscopy. Conservative management is usually successful. Surgical

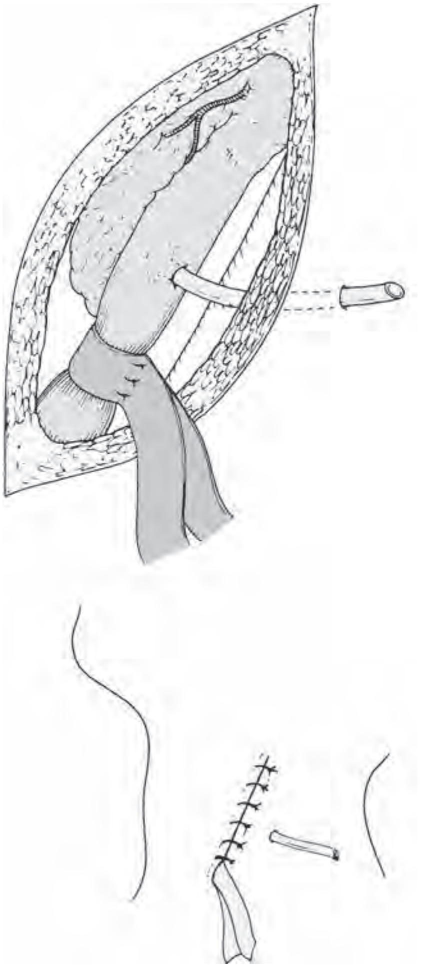


Figure 22.31 Technique of T-tube temporary total diverting cervical oesophagostomy with banding of the oesophagus for high thoracic injuries. The long limbs of the band used to occlude the oesophageal lumen are brought out through the lower end of the incision. (From Hennessey, Cuschieri, *Surgery of the Oesophagus.* Oxford, UK: Butterworth Heineman, 1992, with permission.)

treatment is only indicated for those patients who continue to bleed or in whom the haemorrhage recurs after the above measures. It consists of suture ligation of the bleeding mucosal tears through a generous gastrostomy. Percutaneous embolization of the left gastric artery may be used in poor risk patients such as cirrhotic individuals.

Intramural haematoma

This is very rare entity. It may occur either spontaneously or following transoesophageal echo and cardiac ablation in patients having anticoagulation therapy. In spontaneous cases, it arises from an oesophageal mucosal tear associated with submucosal bleeding with dissection of this plane by the expanding intramural haematoma. Patients give a history of gagging or choking while eating, followed by sharp mid-epigastric/lower retrosternal pain radiating to the back and associated with haematemesis. The patient subsequently develops dysphagia. The diagnosis is usually made on CT imaging and history. The condition is self-limiting.

Atrio-oesophageal fistula

This is a rare condition that typically happens 2 weeks following cardiac interventions. In those cases the majority of patients present with severe sepsis, cerebral abscesses and diminished level of consciousness that requires admission to the intensive care unit. A high level of suspicion is required to reach the diagnosis of those patients. Brain CT is diagnostic, whereas imaging of the chest rarely demonstrates the fistula between the heart and oesophagus. This condition is fatal. In addition to supportive measures, reported cases who survived the condition underwent surgical intervention to repair the heart and oesophagus.

Benign oesophageal strictures

Aetiology

- Gastro-oesophageal reflux
- Postinflammatory conditions
- Post radiation
- Postendoscopic mucosal resection and HALO
- Caustic agents
- Connective tissue disorders, e.g. scleroderma.

Gastro-oesophageal reflux is the most common cause of benign oesophageal strictures. The stricture is caused by chronic fibrosis and scarring in response to chronic reflux injury in the lower oesophagus. Reflux and Barrett's strictures usually occur at the GOJ.

HIV-related benign oesophageal strictures have been reported secondary to idiopathic oesophageal ulceration, cytomegalovirus (CMV) oesophagitis, herpetic oesophagitis and increased sensitivity to radiation therapy. Despite the extensive and deep nature of CMV ulceration, stricture formation is uncommon. There have been anecdotal reports of oesophageal strictures secondary to CMV infection in HIV patients. Oesophageal stricture has been reported during active CMV ulceration and subsequent to successful treatment.

Benign oesophageal strictures occur following radiation therapy to the chest and mediastinum. Oesophageal strictures may also result from chemoradiotherapy treatment of squamous cell carcinoma of the oesophagus. It is crucial in those cases to differentiate between fibrotic benign strictures from recurrent cases and those with residual disease. Also, benign stricture may follow endoscopic therapy of early oesophageal cancer lesions in Barrett's segment such as endoscopic mucosal resection, photodynamic therapy and thermal ablation techniques. Resection of at least 50% of the oesophageal mucosal circumference is strongly associated with stricture formation. Patients with strong histories of tobacco use may also be more likely to develop oesophageal strictures following endoscopic mucosal resections. Again, exclusion of cancerous lesions is mandatory. Benign strictures of the oesophagus occur following oesophageal variceal sclerotherapy in up to 30% of patients and less commonly following oesophageal variceal ligation.

Another cause of benign strictures is ingestion of caustic agents. Caustic stricture formation occurs in up to 30% of patients within 8 weeks and persists. These strictures are often long and may involve the whole length of the oesophagus.

Medication such as alendronate, tetracycline, doxycycline, mycophenolate mofetil, homeopathic pills and NSAIDs can cause oesophageal injury, ulceration and strictures. Rarely, connective tissue disorders such as scleroderma are associated with benign oesophageal strictures.

Clinical picture

Dysphagia is the dominant symptom for oesophageal strictures. The characteristics of dysphagia should be evaluated in terms of onset, duration and the amount of food and nutrition that the patient is able to tolerate. History should be directed to elicit the effects of dysphagia on nutritional status and body weight, and whether the patient has symptoms to suggest respiratory aspiration. History should also seek causes of stricture such as reflux symptoms, endoscopic therapy, radiotherapy, medication and other cause of oesophagitis. An accurate diagnosis is essential, as it affects the treatment options for the oesophageal stricture itself and any underlying disease.

A barium swallow is the initial diagnostic test to establish the length and width of the stricture and the presence of hiatus hernia. It also provides useful information for planning the endoscopy and dilatation. Endoscopic visualization with biopsy and brush cytology should establish the nature of the stricture. The flexible endoscope can also be used for dilatation of the stricture in order to obtain more representative biopsies. A normal barium swallow and endoscopy should rule out a stricture as a cause of dysphagia and in these circumstances a motility disorder of the oesophagus should be considered.

Management

The principles for the management of patients with benign oesophageal strictures are (1) accurate diagnosis of its benign nature and exclusion of malignancy, (2) dilatation to obviate dysphagia and (3) in cases of recurrent symptoms a long-term strategy should be planned.

Dilatation

The principal aim of dilatation is to stretch the fibrous component of the stricture gradually to achieve an oesophageal lumen of 10–15 mm in diameter. The endpoint of dilatation depends on the initial size of the stricture, the general condition of the patient and the predisposing factors. Most strictures can be dilated, but severe long and tortuous strictures are difficult, carry an increased risk of perforation and recur following dilatation. The complications of dilatation include haemorrhage, and perforation. Various types of dilating systems are currently available:

Wire-guided dilators such as the Savery/Gillard dilators, Puestow, Celestin and Gruntzig are useful in all types of strictures. They are more rigid and are less likely to bend or coil. They have a central core so that they can be passed over a preinserted guide wire. Some consist of metal olives (Eder/Puestow) and some are more substantial graduated polyvinyl dilators (Savery/Gillard, Celestin). The guide wire should have a flexible tip and can be inserted either endoscopically or using fluoroscopy. Hydrophilic guide wires tend to be less rigid and more expensive but easier to pass through long and tortuous strictures. When the guide wire is inserted endoscopically it is safer to check the position of the wire by fluoroscopy. When the guide wire is in position the endoscope can be withdrawn and the dilators passed over the guide wire across the strictured part of the oesophagus. Advancement of the dilator should be coupled with rotation of the dilator around the guide wire. When undue resistance is encountered, dilatation should be halted. For severe strictures the dilatation requires to be gradual and the procedure may necessitate two or more sessions with rest intervals of 1–2 weeks between sessions. During each session, dilatation should not exceed a 6–8 mm increase in diameter. If available, routine fluoroscopy should be used for all wire-guided dilatations.

Hydrostatic polyurethane balloon dilators are of two types: those used via the endoscope and dilators over a guide wire under fluoroscopy. The balloons are hydrostatically pressurized to inflate to a maximum diameter. Some versions produce a variable balloon maximum diameter dependent on the hydrostatic pressure. The balloon catheter tip is flexible and atraumatic and can negotiate tortuous or narrow strictures. Hydrostatic balloons are most popular, largely safe and can be used progressively to dilate long strictures.

Once adequate dilatation is achieved, patients should be advised on lifestyle adjustment and should be prescribed PPIs. Patients with reflux symptoms may require antireflux procedures. The majority of patients achieve a satisfactory freedom from dysphagia using this approach. The rest will require frequent dilatations. A single dilatation usually obviates all oesophageal rings and webs; however, medication-and caustic-induced strictures and radiation strictures may require repeat dilatations.

Stent insertion

For patients who require repeated dilatations, removable fully covered self-expandable metal or plastic stents can be inserted to provide long-term alleviation of dysphagia. The success rate in complete remission of dysphagia is 30–50%. Those stents are usually endoscopically removed after 4 weeks. Biodegradable stents have been increasingly used recently to treat benign oesophageal strictures. They have a lower reintervention rate and may potentially reduce complications during stent removal. Complications of stent insertion are haemorrhage and perforation. If the strictures are resistant to stent treatment and the patients are fit, oesophagectomy and replacement surgery should be considered.

Surgical treatment

Indications for surgery are fit young patients in whom repeated dilatation and stent intubations have failed to alleviate dysphagia. Also, surgery is indicated in impassable strictures associated with Barrett's oesophagus as malignancy may not be excluded with certainty.

All young patients with peptic strictures that respond to dilatation and PPIs should be considered for standard *antireflux surgery*. If there is significant oesophageal shortening a *Collis gastroplasty* is indicated to create a neo-oesophagus around which a wrap is formed.

Resection of refractory benign oesophageal strictures is the next therapeutic option. Lower oesophageal segment/gastro-oesophageal stricture can be replaced by *interposed isoperistaltic jejunum*. For long strictures, *subtotal oesophagectomy* with replacement by stomach or colon is reserved. The best long-term results are probably achieved by a vagus-sparing oesophagectomy with an interposed isoperistaltic jejunum or left colon placed in the posterior mediastinum. The use of a gastric pull-through would necessarily incur division of the vagus but would make the operation technically easier and less demanding for the patient.

Oesophageal cancer

Epidemiology

Oesophageal adenocarcinoma has been rapidly increasing in Western countries during the past half century, especially in white males. In the UK, USA and Australia, the incidence of oesophageal adenocarcinoma among white men was 0.5–0.9 per 100 000 population in the 1970s but has increased to 3.2–4 per 100 000 over the following two decades. In these countries the ratio of squamous cancer to adenocarcinoma has decreased significantly, with the dominant cancer in Western countries being oesophageal adenocarcinoma since the 1980s. Among Asian countries, the incidence of oesophageal adenocarcinoma is quite low and the incidence of oesophageal squamous carcinoma is higher, especially in East Asia. Oesophageal squamous cell carcinoma is the dominant type of oesophageal cancer in China, Hong Kong, Taiwan, Korea and Japan.

The increase in oesophageal adenocarcinoma seems to be correlated with the increased prevalence of GORD and being overweight. The risk associated with GORD is related to Barrett's metaplasia. It was found that the pattern of ethnic differences in Barrett's oesophagus is quite similar to that found in the relative incidence of oesophageal adenocarcinoma in epidemiological studies. It was also shown that a raised body mass index (BMI) is a risk factor for oesophageal adenocarcinoma, independently of the occurrence of gastro-oesophageal reflux. Barrett's oesophagus and oesophageal adenocarcinoma seem to be in parallel with the development of socioeconomic status and, geographical differences in the epidemiology of oesophageal carcinoma seem to be correlated with the socioeconomic status.

Aetiology

Gastro-oesophageal reflux

In a meta-analysis of the association of adenocarcinoma with symptoms of GORD, frequent symptoms are strongly associated with oesophageal adenocarcinoma. At least weekly symptoms of GORD increases the odds of oesophageal adenocarcinoma fivefold and daily symptoms increase the odds sevenfold, each compared with individuals without symptoms or less frequent symptoms.

Obesity

A meta-analysis of 14 studies has found that oesophageal adenocarcinoma risk increased two-to threefold in overweight and obese patients. It also observed a dose-dependent relationship, with the risk being higher in obese than in overweight people. Several additional population-based cohort studies of obesity and oesophageal adenocarcinoma risk also showed two- to threefold increased risk and a dose-dependent relationship.

Also obesity predisposes to hiatus hernia and reflux, which contributes to the increased risk of oesophageal adenocarcinoma. There is increased risk only found in obese women (BMI >30), whereas in men it was observed in both overweight (BMI 25–29.9) and obese (BMI >30). In the Million Women Study in the UK, 50% of cases of oesophageal adenocarcinoma in postmenopausal women were attributed to obesity. It was also shown that the male pattern of abdominal obesity (central and retroperitoneal) is more likely to be associated with malignant transformation.

Smoking

In another meta-analysis of 33 studies, compared with never smokers the pooled relative risk was 1.76 for ever smokers, 2.32 for current smokers and 1.62 for ex-smokers. There was a direct association with those with more than 20 cigarettes a day and duration for more than 70 years of cigarette consumption.

Alcohol consumption

The increased risk of oesophageal cancer associated with alcohol use is perhaps limited to oesophageal squamous cell carcinoma. Excessive amounts of alcohol (three or more drinks per day) has almost universally been associated with an elevated risk of oesophageal squamous cell carcinoma, typically increasing risk by three- to fivefold. In contrast, there is little evidence for an association between alcohol drinking and oesophageal adenocarcinoma. The exact mechanism of carcinogenicity of alcohol is not known because alcohol itself does not bind DNA, is not mutagenic and does not cause cancer in animals. However, mechanisms have been suggested, including its conversion to acetaldehyde acting as a solvent for other carcinogens and causing nutritional deficiency.

Helicobacter pylori

Meta-analysis showed that *H. pylori* colonization of the stomach is associated with an early 50% reduction in the risk of oesophageal adenocarcinoma. *H. pylori* may decrease the risk of oesophageal carcinoma by reducing gastric acid production and hence reducing acid reflux from the stomach to the oesophagus. It may also reduce oesophageal adenocarcinoma risk by decreasing the production of ghrelin, a hormone that is mostly produced in the stomach and stimulates appetite. A reduction in the level of ghrelin may lead to lower rates of obesity, an important risk factor for oesophageal adenocarcinoma. In contrast to its association with oesophageal adenocarcinoma, *H. pylori* has not shown a consistent association with oesophageal squamous cell carcinoma. Three meta-analyses on the association of *H. pylori* with squamous cell carcinoma found no overall association, with the summary odds ratio close to unity. In a case-controlled study, the role of *H. pylori* infection and atrophic gastritis was examined in cardia cancer. The study suggested there were two categories of cardia cancer, one associated with *H. pylori* atrophic gastritis resembling non-cardia gastric cancer and the other associated with non-atrophic gastric mucosa resembling oesophageal adenocarcinoma.

Non-steroidal anti-inflammatory drugs

A meta-analysis of nine studies concluded that aspirin and other NSAIDs reduce the risk of oesophageal cancer in a dose–response manner. More frequent use was associated with lower risk, with an overall risk reduction of approximately 40%. This meta-analysis showed similar inverse association with oesophageal squamous cell carcinoma and oesophageal adenocarcinoma. This association may be due to reverse causation: people with a history of upper gastrointestinal symptoms who are also at high risk of cancer may limit the use of aspirin and other NSAIDs. However, some studies stratified their results by having or not having upper gastrointestinal disorder and did not find a difference in the results in these two groups of patients. If the associations are causal, aspirin and NSAIDs may reduce risk by reducing inflammation and affecting the inflammation–metaplasia–cancer sequence at the earliest stage.

Achalasia

Most case studies have reported oesophageal carcinoma in 3–7% of achalasia patients, which is higher than the rates seen in normal populations. Follow-up studies of achalasia patients have consistently shown a substantial increase in the risk of oesophageal carcinoma. One of the largest studies in Sweden showed a 10-fold increased risk in both oesophageal squamous cell carcinoma and oesophageal adenocarcinoma in achalasia patients compared with the rest of the population.

Eating pattern

There is a decreased risk of oesophageal squamous cell carcinoma but not oesophageal adenocarcinoma with higher intake of both fruit and vegetables. On the other hand, eating pickled vegetables was once very common in high-risk areas of China and was thought to be a major risk factor for oesophageal carcinoma in those areas. The results of epidemiological studies have been inconsistent, where some case-controlled studies have shown an association between pickled vegetable intake and oesophageal carcinoma, typically with increased risk of two- to threefold. Other case–control or cohort studies have shown no association.

Poor oral hygiene and tooth loss

Case–control studies showed that tooth loss was associated with a two- to threefold increase in oesophageal carcinoma. Nearly all studies of the association between poor oral health and oesophageal carcinoma risk has been done in areas of oesophageal squamous cell carcinoma. To date there have been no studies specifically evaluating the association between poor oral health and oesophageal adenocarcinoma.

Low socioeconomic status

It has been long known that oesophageal carcinoma is a disease of the poor and socially disadvantaged. Large numbers of epidemiological studies have confirmed that oesophageal carcinoma risk is higher in the population with lower socioeconomic status using different designs (case–control, cohort, comparison of incident cases with the general population), and different socioeconomic class indices and from all parts of the world. The finding that oesophageal carcinoma is more common in lower socioeconomic status is universal. Most studies reported an increased risk of two- to fourfold among those with lower than those with higher socioeconomic status. The majority of studies of socioeconomic status and oesophageal carcinoma have been conducted in populations with a high risk of oesophageal squamous cell carcinoma. There are few data concerning risk association with oesophageal adenocarcinoma, although rich socioeconomic status may link to oesophageal adenocarcinoma.

Summary

Three important risk factors have been identified for oesophageal adenocarcinoma: gastro-oesophageal acid reflux, obesity and smoking. Absence of *H. pylori* in the stomach is also becoming an increasingly recognized risk factor.

Excessive use of alcohol, tobacco use, low intake of fruit and vegetables and low socioeconomic status are risk factors for oesophageal squamous cell carcinoma. Achalasia may predispose people to high risk of oesophageal squamous cell carcinoma.

Clinical presentation

Dysphagia is one of the classical symptoms of oesophageal cancer. It is reported to be present in about 40–50% of patients with a mean duration of about 2 months. It starts with solid food and progresses towards soft food and fluids. Dysphagia usually presents late in the natural history of the disease and becomes severe only when more than 60% of the oesophageal circumference is infiltrated with cancer. The majority of patients who present with dysphagia also have weight loss due to inadequate oral intake and as a consequence of advanced stage of the disease. Any patient with dysphagia should be investigated with upper gastrointestinal endoscopy regardless of age.

There are increasing numbers of patients with early disease who are diagnosed during surveillance for Barrett's oesophagus or investigations for long history of GORD, accounting for about 20–25% of patients at the time of presentation. About 25–30% of patients present with anaemia, chest/abdominal discomfort and bleeding.

Clinical factors that indicate an advanced stage of carcinoma and exclude surgery with curative intent are recurrent nerve paralysis, Horner syndrome, persistent spinal pain, paralysis of the diaphragm, fistula formation and malignant pleural effusion.

Investigations to stage oesophageal cancer

The aim of investigations for patients presenting with the above symptoms is to confirm the nature of the disease and determine the stage of cancer. A multimodality approach including upper gastrointestinal endoscopy, CT, positron emission tomography (PET), EUS and laparoscopy is recommended.

Upper gastrointestinal endoscopy

Upper gastrointestinal endoscopy is the gold standard first investigation to diagnose oesophageal cancer. It allows biopsy tissue to be obtained for diagnosis and the description of the level and morphology of the tumour. Accurate documentation of the length and the level of the tumour are crucial as they will influence management options. Skip cancer lesions can be also seen and biopsied. The appearance of small early lesions will signal the need for further evaluation with a view to endoscopic curative management rather than the surgical option. The ability to pass the endoscope to the stomach should also be reported, as it indicates the ability of the patient to maintain oral intake.

Failure to diagnose malignancy in the first endoscopy is in the range of 10% and about 10–20% of patients require a second endoscopy. If the histology does not confirm oesophageal cancer but the endoscopist has a high suspicion of cancer from the appearance of the lesion, endoscopy should be repeated and further tissue should be sent for histological evaluation. Multiple four-quadrant biopsies of Barrett's lesions at 2 cm intervals along its entire length have been shown to increase diagnostic accuracy and allow differentiation of high-grade dysplasia from adenocarcinoma. For patients with Barrett's oesophagus, continuing treatment with antacid medications can decrease inflammation, making targeted biopsies and histological examination easier.

Chromoendoscopy and high-resolution endoscopy have been introduced to increase the diagnostic accuracy of high-grade dysplasia and early cancer lesions. Lugol's iodine is used for dysplastic and malignant squamous mucosa whereas acetic acid chromatography enhances the detection of occult neoplasia in Barrett's oesophagus. Imaging modalities such as narrow-band and zoom (magnifying) endoscopy are used in selected centres to increase diagnostic accuracy of early lesions. Nevertheless those techniques are not currently in routine clinical practice.

Computed tomography

CT of the chest, abdomen and pelvis is the investigation of choice after confirming the nature of oesophageal lesions on endoscopic biopsies. It is widely available and indicates the local stage and the presence of metastasis. A CT scan should be performed with intravenous contrast (unless contraindicated) and distension of the oesophagus and stomach with about 1 L of water (as long as patients can tolerate this amount) and sometimes gas-forming granules. The use of thin slices (2 mm) and multiplanar reformatted images provides accurate information about the extent and possible invasion of the tumour to adjacent structures such as the diaphragm and aorta.

CT is generally not helpful for T1 disease but provides important information for more advanced stages. The accuracy of CT on staging oesophageal cancer shows a sensitivity and specificity of about 84% for T staging, a sensitivity of 50% and specificity of 85% for N status and a sensitivity of 52% and specificity of 91% for metastatic disease.

^{18}F-Fludeoxyglucose positron emission tomography

Recently, PET and in particular PET-CT are emerging as valuable tools in the staging and in providing prognostic information such as the response to neoadjuvant chemotherapy. Meta-analysis on the yield of CT scan on staging oesophageal cancer showed a sensitivity and specificity of about 87% for T staging, a sensitivity of 50% and specificity of 83% for N status and a sensitivity of 71% and specificity of 93% for metastatic disease. In one study, patients with no metastasis on CT scan, ^{18}F-fludeoxyglucose (FDG)-PET identified distant disease in 4.8% of patients. The distant metastasis in PET should be carefully evaluated and may be confirmed by tissue biopsy before excluding patients from curative surgery, as at least 3.7% are false positives.

Endoscopic ultrasound

This modality has the highest accuracy in evaluating the local stage of the disease. EUS has gained popularity over the last 10 years as the modality of choice for pretherapeutic assessment of depth of tumour infiltration into the oesophageal wall and lymph node status. EUS was confirmed to be reasonably accurate in the identification of T3 tumours (83%). T2 tumours were correctly identified by EUS in only 42% of patients. Even less accurate was the identification of T1 or *in situ* tumours. T1 lesions were correctly classified in 29% of patients, but *in situ* tumours cannot be correctly identified preoperatively.

EUS has a specificity of 70% and a sensitivity of 80% in the diagnosis of regional lymph node metastasis. Although EUS has less overall accuracy for T1 disease, it is important for detecting nodal metastasis when planning endoscopic curative therapies. Endosonographic features predictive of malignancy in increasing order of importance are echo-poor (hypoechoic) structure, sharply demarcated borders, rounded contour and size greater than 10 mm. Collectively, the EUS features produce an additive effect with respect to accuracy in the prediction of malignant lymph node involvement; malignancy could be predicted with 100% accuracy when all four features are present. A careful and systematic approach to the endosonographic assessment of lymph node metastasis can improve staging accuracy.

One of the advantages of EUS is the ability to obtain fine-needle aspiration cytology to confirm local lymph node involvement. EUS can also determine the exact longitudinal extent of the disease, because of its ability to diagnose submucosal extension. The EUS report includes the start and end of the tumour and any involvement of the stomach as well as the level of landmark structures such as carina, crura of the diaphragm and GOJ. This is important in deciding the surgical approach and the extent of resection.

The accuracy of EUS decreases dramatically after neoadjuvant chemotherapy. It also has a limited value when malignant strictures do not allow the passage of the endoscopy for complete assessment.

Staging laparoscopy

Staging laparoscopy should be performed with all lower oesophageal and GOJ tumours to detect peritoneal and small liver deposits that are too small for the resolution of CT and PET imaging. It is also important to inspect the cardia and lesser curvature for tumour extension into the stomach. Serosal gastric extension can sometimes be seen during laparoscopy. Peritoneal lavage is routinely done during laparoscopy for lower oesophageal and gastro-oesophageal junctional tumours. Additional information from staging laparoscopy that influences the treatment decision is obtained in about 15% of patients. Many surgeons prefer to repeat the upper gastrointestinal endoscopy themselves at the time of staging laparoscopy to inspect the tumour and plan the surgical approach.

Bronchoscopy

This is important for tumours at or above the level of the carina that are adherent to the bronchial tree. The examination may be supplemented with endobronchial ultrasound, which also allows fine-needle aspiration of lymph nodes along the trachea that cannot be approached via EUS.

Stages of oesophageal cancer (TNM version 7)

This classification applies to carcinoma and includes adenocarcinoma of the oesophagogastric junction. There should be histological confirmation of the disease and division of cases by topographic localization and histological type. A tumour the epicentre of which is within 5 cm of the oesophagogastric junction and which also extends into the oesophagus is classified and staged using the oesophageal scheme. Tumours with an epicentre in the stomach greater than 5 cm from the oesophagogastric junction or those within 5 cm of the oesophagogastric junction without extension in the oesophagus are classified and staged using the gastric carcinoma scheme. The procedures for assessing T, N and M categories are physical examination, imaging, endoscopy (including bronchoscopy) and/or surgical exploration.

Anatomical subsites

1. Cervical oesophagus: this commences at the lower border of the cricoid cartilage and ends at the thoracic inlet (suprasternal notch), approximately 18 cm from the upper incisor teeth.
2. Intrathoracic oesophagus
   a. The upper thoracic portion extends from the thoracic inlet to the level of the tracheal bifurcation, approximately 24 cm from the upper incisor teeth.
   b. The mid-thoracic portion is the proximal half of the oesophagus between the tracheal bifurcation and the oesophagogastric junction. The lower level is approximately 32 cm from the upper incisor teeth.
   c. The lower thoracic portion, approximately 8 cm in length (includes abdominal oesophagus), is the distal half of the oesophagus between the tracheal bifurcation and the oesophagogastric junction. The lower level is approximately 40 cm from the upper incisor teeth.
3. Oesophagogastric junction.

Regional lymph nodes

The regional lymph nodes, irrespective of the site of the primary tumour, are those in the oesophageal drainage area including the coeliac axis nodes and the paraoesophageal nodes in the neck, but not the supraclavicular nodes.

TNM clinical classification

T – Primary tumour

- TX: Primary tumour cannot be assessed
- T0: No evidence of primary tumour
- Tis: Carcinoma *in situ*/high-grade dysplasia
- T1: Tumour invades lamina propria, muscularis mucosae or submucosa
  - T1a: Tumour invades lamina propria or muscularis mucosae
  - T1b: Tumour invades submucosa
- T2: Tumour invades muscularis propria
- T3: Tumour invades adventitia
- T4: Tumour invades adjacent structures
  - T4a: Tumour invades pleura, pericardium, or diaphragm
  - T4b: Tumour invades other adjacent structures such as aorta, vertebral body or trachea

N – Regional lymph nodes

- NX: Regional lymph nodes cannot be assessed
- N0: No regional lymph node metastasis
- N1: Metastasis in one to two regional lymph nodes
- N2: Metastasis in three to six regional lymph nodes
- N3: Metastasis in seven or more regional lymph nodes

M – Distant metastasis

- M0: No distant metastasis
- M1: Distant metastasis

pTNM pathological classification

The pT and pN categories correspond to the T and N categories.

- pN0: Histological examination of a regional lymphadenectomy specimen will originally include six or more lymph nodes. If the lymph nodes are negative, but the number ordinarily examined is not met, classify as pN0.
- pM1: Distant metastasis microscopically confirmed

G – Histopathological grading

- GX: Grade of differentiation cannot be assessed
- G1: Well differentiated
- G2: Moderately differentiated
- G3: Poorly differentiated
- G4: Undifferentiated

Stage grouping of carcinomas of the oesophagus and oesophagogastric junction:

- Stage 0: TisN0M0
- Stage 1A: T1N0M0
- Stage 1B: T2N0M0
- Stage IIA: T3N0M0

- Stage IIB: T1,T2N1M0
- Stage IIIA: T4aN0M0
  - T3N1M0
  - T1,T2N2M0
- Stage IIIB: T3N2M0
- Stage IIIC: T4aN1,N2M0
  - T4b: Any NM0
  - Any TN3M0
- Stage IV: Any T, Any NM1

Summary: oesophagus (includes oesophagogastric junction)

- T1: Lamina propria (T1a), submucosa (T1b)
- T2: Muscularis propria
- T3: Adventitia
- T4a: Pleura, pericardium, diaphragm
- T4b: Aorta, vertebral body, trachea
- N1: one or two nodes
- N2: three to six nodes
- N3: seven or more regional
- M1: Distant metastasis

Adenocarcinoma at the oesophagogastric junction

Siewert and Stein proposed the classification of adenocarcinoma of the GOJ in 1996, which was based on topographic anatomical criteria. Adenocarcinoma of the GOJ is divided into three types, each of which had different characteristics, thereby influencing the selection of the surgical strategy:

- Type I: tumours in which the centre is located 1–5 cm above the GOJ, regardless of invasion to the GOJ.
- Type II: tumours invading the GOJ, in which the centre is located between 1 cm above and 2 cm below the GOJ.
- Type III: tumours invading the GOJ, in which the centre is located 2–5 cm below the GOJ.

This classification was approved at the consensus conference of the International Gastric Cancer Association (IGCA) and the International Society for Diseases of the Oesophagus (ISDE), and has been accepted and is now used worldwide. The Siewert subtype should be determined prospectively, based on the findings of endoscopy, contrast radiography and CT. A definition of anatomical cardia is determined by the findings of endoscopy.

Theoretically, type I tumours arise from the oesophageal glandular epithelium or Barrett's oesophagus. The prevalence of Barrett's oesophagus in patients with type I tumours is higher than that in patients with type II/III tumours, in both Western and Eastern countries. Type II tumours are true adenocarcinoma of the oesophagogastric junction arising from the junctional epithelium; however, some type II tumours can arise from the same origin as type I tumours, and some can arise from the same origin as type III tumours. Many previous studies have demonstrated that the characteristics of type II tumours are more like those of type III tumours than those of type I tumours, thus indicating that the origin of type II tumours is similar to that of type III tumours. Type III tumours arise from the gastric mucosa, and this origin might be associated with *H. pylori* and atrophic gastritis. Tumours whose centre is located 2–5 cm below the GOJ are classified as non-junctional or true gastric cancers when the tumours do not invade the GOJ. The tumour classification changes to AEG type III when they invade the GOJ by horizontal progression. Therefore, subcardial gastric cancers are classified as type III tumours when they are enlarged. Type III tumours should be treated as gastric cancers invading the GOJ, considering the origin of the tumours. In summary, adenocarcinoma of the oesophagogastric junction may contain two distinct aetiologies. It is often difficult to determine the tumour origin, especially in advanced cases.

Management of disease confined to the mucosa

Vascular invasion and lymph node metastasis are very uncommon in intramucosal cancer (T1a), whereas 25% of patients with submucosal cancer (T1b) have lymph node metastasis. Therefore high-grade dysplasia and intramucosal cancer are suitable candidates for endoscopic curative interventions. There are several options to treat those early lesions. The multidisciplinary team should determine the management strategy depending on local expertise of the surgical and medical teams and on the availability of different endoscopic technologies.

Surgical resection has been the curative option in the majority of centres for patients with high-grade dysplasia and carcinoma *in situ*. The rationale behind the surgical approach is (1) the curative nature of the interventions as no oesophageal cancer-related death has been reported after surgery; (2) the demonstrated safety of surgery in high-volume centres with a postoperative mortality approaching zero in large series; (3) the adequacy of the non-radical approach, such as transhiatal, vagal sparing and minimal invasive oesophagectomy, with its reduced impact on postoperative adverse events and quality of life; and (4) the discovery of incidental intramucosal or invasive cancer in resected specimens that were not diagnosed preoperatively in the range of 15–40% in large series. However, surgery is associated with significant morbidity, mortality in high-risk patients and change in the quality of life.

Recently, with the development of different endoscopic therapies, endoscopic interventions are gaining popularity as an alternative favourable option to surgery. The combination of mucosal resection and radiofrequency ablation are the current popular strategy. The rationale behind this approach is (1) adequate pathological examination with mucosal resection to exclude submucosal invasive cancer; (2) the low rate of 5% or less of missing invasive cancer with the advances in diagnostic endoscopic techniques; (3) safety of the techniques even in patients with significant comorbidity; and (4) avoiding side effects of surgery in terms of postoperative complications and the effect on quality of life. However, patients undergoing definitive endoscopic therapy require a careful follow-up programme with frequent endoscopic examinations.

Endoscopic interventions are divided into mucosal ablative therapy and endoscopic mucosal resection. Ablative therapies destroy the metaplastic epithelium allowing replacement with neosquamous epithelium. The main ablation modalities are radiofrequency ablation and photodynamic therapy. In radiofrequency ablation, a high-power short burst of energy is applied to the columnar epithelium via direct contact. A dose of 10–12 Js/cm^2 removes all epithelium without damage to the underlying submucosa, resulting in a lower incidence of treatment-related stricture and improved rates of ablation of dysplasia and Barrett's oesophagus. In a multicentre sham-controlled randomized trial, 127 patients with dysplastic Barrett's oesophagus were randomized to receive either radiofrequency ablation (ablation group) or a sham procedure (control group). In the intention-to-treat analyses, among patients with low-grade dysplasia, complete eradication of dysplasia occurred in 90% of those in the ablation group, compared with 23% in the control group. Among patients with high-grade dysplasia, complete eradication occurred in 81% in the ablation group, compared with 19% in the control group. Overall, 77% of patients in the ablation group had complete eradication of intestinal metaplasia, compared with 2.3% in the control group. Patients in the ablation group had less disease progression (3.6% vs 16.3%) and fewer cancers (1.2% vs 9.3%). Patients reported having more chest pain after the ablation procedure than after the sham procedure. In the ablation group, one patient had upper gastrointestinal haemorrhage, and five patients (6.0%) had oesophageal stricture. This study has clearly documented the value and safety of radiofrequency ablation in the management of dysplastic oesophagus. This modality has replaced photodynamic therapy, which suffers from a high prevalence of incomplete ablation, 30% prevalence of strictures and lack of availability in most centres.

A limitation of all ablative therapies is the lack of specimens for histological examination. With endoscopic mucosal resection, this limitation is overcome. Owing to the inability to see high-grade dysplastic lesions for mucosal resection and the frequency of multifocal carcinoma within dysplastic Barrett's oesophagus, a concomitant ablative procedure of Barrett's oesophagus is required to ensure complete eradication of disease. Endoscopic mucosal resection should generally be limited to less than 5 cm in diameter and used primarily for resection of nodules associated with Barrett's oesophagus rather than for complete Barrett's eradication. In a study of 100 consecutive patients with low-risk adenocarcinoma of the oesophagus who underwent endoscopic resection, no major complications and only 11 minor cases of bleeding were reported. Complete local remission was achieved in 99 of the 100 patients after 1.9 months with a maximum of three resections. During a mean follow-up period of 36.7 months, recurrent or metachronous carcinomas were found in 11% of the patients, but successful repeat treatment with endoscopic resection was possible in all of these cases. The calculated 5 year survival rate was 98%. Two patients died of other causes.

There are no prospective trials comparing outcomes of endoscopic and surgical interventions for Barrett's oesophagus and high-grade dysplasia, although a number of retrospective studies have been reported. Those studies do not show a survival difference for either surgery or ablation with or without mucosal resection, although early mortality was higher in the surgical group. There was significant treatment failure observed in the endoscopic intervention groups as 10–20% of patients develop new or metachronous cancer. There was no cancer recurrence in the surgical group.

Surgical management of oesophageal cancer

Preparation of patients for oesophageal surgery is crucial. This should entail a discussion by the multidisciplinary oesophageal cancer team for accurate cancer staging and evaluation of fitness for surgery. Optimization of cardiorespiratory and nutritional status is essential to minimize postoperative complications. Stopping smoking, regular exercise and adequate nutrition should be routine advice for patients undergoing oesophageal surgery.

Different surgical approaches

The following definitions describe access to oesophagectomy:

- *Transthoracic oesophagectomy:* open resection of the oesophagus employing thoracotomy, including all single-, two- and three-site approaches utilizing either a right or left thoracotomy or thoracoabdominal incisions.
- *Transhiatal oesophagectomy*: open resection of the oesophagus performed through the oesophageal hiatus and thoracic inlet without open thoracotomy.
- *Minimally invasive oesophagectomy*: minimal access approach via laparoscopy and/or thoracoscopy. A hybrid procedure is often performed combining a minimal access approach with laparotomy or thoracotomy.

Extent of lymphadenectomy

This describes the field of lymphadenectomy performed in the abdomen, chest or the neck as follows:

- Radical lymphadenectomy procedures
  - *Three-field lymphadenectomy* with lymph node resection encompassing upper abdominal (D2), inferior middle and superior mediastinal lymph nodes, and inferior cervical nodes.
  - *Two-field lymphadenectomy*, which is the same as the above with the exception of removal of the cervical nodes.
  - *Infracarinal two-field lymphadenectomy* which omits both cervical lymph nodes and a formal superior mediastinal nodal dissection.
- Non-radical lymphadenectomy is a conservative lymphadenectomy in which only the nodes in direct proximity to the tumour, the oesophagus and upper stomach are removed.

It is important for surgeons not to mix the approach with the field of lymphadenectomy. Ivor Lewis oesophagectomy entails laparotomy and right thoracotomy. This may or may not entail radical lymphadenectomy in the form of abdominal and infracarinal two-field lymphadenectomy.

Rationale for lymphadenectomy

Prevelance of lymph node metastasis with oesophageal cancer

Lymph node involvement is very common with oesophageal cancer even with early disease. As tumour depth increases the extent of lymph node involvement increases in terms of both the number of nodes involved and the involvement of distant nodes. Lymph node metastases have been found in up to 35% of T1b tumours and in 78–85% of T3 tumours. Lymph node involvement may occur in the abdominal, thoracic or cervical nodes regardless of whether the primary tumour is located at the GOJ, distal, mid- or upper oesophagus. Metastatic nodes in the cervical lymph nodes have been found in 17% of GOJ tumours and 23% of distal third tumours.

Accurate staging

The Surveillance, Epidemiology and End-Results (SEER) database study of 2597 patients showed that N0 had fewer lymph nodes identified than those classified as N. Other investigators have reported a strong correlation between the number of lymph nodes examined and the number involved. For example, patients who were N0 had a mean lymph node count of 14.2 ± 7.1, whereas the lymph node count was 18.0 ± 9.3 in patients found to have lymph node metastases. In another study of 366 patients, those who had 15 lymph nodes examined were more likely to have positive lymph node metastases than patients who had less than 15 nodes examined.

Influence of the extent of lymphadenectomy on survival

Several authors in Western centres have reported case series of radical lymphadenectomy with improved long-term survival (Altorki and Skinner; Lerut *et al.*; Hagen and DeMeester). Also, radical three-field lymphadenectomy is the standard surgical approach in Japan with 5 year survival of 68–54%. A randomized trial of cervical and upper mediastinal lymph node dissection for squamous cell carcinoma of the thoracic oesophagus reported overall survival of 66% at 5 years for the extended dissection with harvest of a mean of 82 nodes vs 48% for standard dissection with a harvest of 43 nodes. In a study examining the effect of one-, two- or three-field lymphadenectomy increased survival was reported with increasing extent of lymph nodes resection. Overall survival at 5 years was 21.2% vs 36.3% vs 53.7% for patients resected with a one-, two- or three-field lymphadenectomy. However, the benefit of more extensive lymphadenectomy was predominantly in the N0 group, suggesting that the benefit is predominantly related to stage migration.

Influence of the number of lymph nodes resected on survival

The SEER 1973–2003 database study examined the relationships between the number of lymph nodes examined and overall survival in a cohort of 5620 patients. On multivariate analysis, total lymph node count (or negative lymph node count, respectively) was an independent prognostic variable. Higher total lymph node count (>30) and negative lymph node count (>15) categories were associated with best overall survival and lowest 90 day mortality. The numeric lymph node effect on overall survival was independent from nodal status or histology. Greater total and negative lymph node counts are associated with longer survival. Although the mechanism remains uncertain in this study, it does not appear to be limited to stage migration. There was an optimum cut-off point for the number of lymph nodes resected above which there was no further improvement in survival.

Another international study of 2303 patients from nine centres who underwent R0 oesophagectomy showed that the number of lymph nodes removed is an independent predictor of survival after oesophagectomy for cancer. To maximize this survival benefit a minimum of 23 regional lymph nodes must be removed.

Influence of lymphadenectomy on local recurrence

Local recurrence has been shown to be very uncommon (<5–10%) following radical three-field lymphadenectomy, which contrasts with a 35% prevalence of local failure following transhiatal oesophagectomy and a 15–20% prevalence after Ivor Lewis resection. These findings may be explained by the removal of unrecognized, often microscopic, disease with more extensive lymph node dissection. Involved nodes that are left behind during simple oesophagectomy are probably the cause of recurrent local disease. Local recurrence is very difficult to treat and affects the quality of life and may also influence long-term survival.

Safety of radical procedures

With centralization of surgical services, the postoperative mortality of radical oesophagectomy has fallen significantly to <5% in Western centres. In some centres, the mortality reaches <1%. High hospital volume drives the process of care and therefore the improvement in the postoperative pathway and in managing postoperative complications results in a significant reduction in mortality in spite of a relatively high postoperative complication rate.

Quality of evidence for comparative studies of transthoracic vs transhiatal oesophagectomy

In a systematic review of the quality of evidence and meta-analysis of outcomes, survival was shown to be equivalent in randomized trials and comparative studies but more advanced cancer and inadequate surgical quality (including inadequate lymph node clearance) of transthoracic oesophagectomy were found to be significant confounding factors. Although transhiatal resection was chosen significantly more frequently for early stage tumours, the transthoracic approach was more often adopted for advanced stage IV tumours. This may be influenced by stage migration since significantly more lymph nodes were harvested from patients in the transthoracic group.

Also, rating the reported quality of surgical resections shows that no study met all minimum surgical quality standards. Only four of the 52 studies presented lymph node data in a format that was suitable for meta-analysis. This is of significant concern as the number of resected lymph nodes has been shown to be a determinant factor for survival in oesophageal cancer. Therefore, evidence from those randomized trials and comparative studies should be viewed with caution.

The extent of radicality of lymphadenectomy

If radical lymphadenectomy provides better survival and lower local recurrence rate, it is expected that three-field lymphadenectomy would have superior results to infracarinal lymphadenectomy. This is particularly with squamous cell carcinoma as 30% of patients will have lymph node metastasis in the neck and around recurrent laryngeal nerves in the superior mediastinum. It is also likely to be the case, although less evidence is available, for adenocarcinoma where 15% of patients have positive cervical lymph nodes. The vast majority of patients have T2 and T3 tumours, with the majority having positive lymph nodes. Some surgeons in Western centres have moved from infracarinal lymphadenectomy to three-field lymphadenectomy.

Transthoracic vs transhiatal oesophagectomy

Meta-analysis of 52 studies, comprising 5905 patients (3389 transthoracic and 2516 transhiatal) showed that transthoracic operations took longer and were associated with a longer length of stay. There was no difference in blood loss. The transthoracic group had significantly more respiratory complications, wound infections and early postoperative mortality, whereas anastomotic leak, anastomotic stricture and recurrent laryngeal nerve palsy rate were more common in the transhiatal group. The lymph node harvest was only reported in four studies and was significantly greater in the transthoracic group by on average only eight lymph nodes. Complete 5 year follow-up was only present in a single paper and using actuarial 5 year survival there was no significant difference found. No study met all minimum surgical quality standards and therefore the evidence of those studies should be viewed with caution.

Technical aspects of oesophagectomy

Stapled vs hand-sewn anastomosis

Meta-analysis of 10 studies comprising 936 patients and comparing hand-sewn with stapled anastomosis showed 10.04% anastomotic leaks in the hand-sewn group and 51 (10.9%) in the stapled anastomosis group with no significant difference between the groups.

Methods of reconstruction

This depends on the location of the tumour and previous surgical procedures. Stomach conduit is the most commonly used. The stomach has a reliable blood supply and its mobilization can be performed with standard lymph node dissection. The reconstruction has a single anastomosis. Disadvantages include the loss of the gastric reservoir and increased risk of acid reflux.

Colonic reconstruction has the potential advantages of providing a peristaltic conduit, preserving the gastric reservoir and, overall, has a lower incidence of postresection oesophageal reflux. Disadvantages include a much longer, more complex operation with three anastomoses, a less predictable blood supply and an increased risk of internal hernias, and a tendency for even well-constructed colonic interpositions to become tortuous and dilated over time.

Roux-en-Y jejunal reconstructions are used when stomach or colon is unavailable and are limited with extent to proximal extension, although optimum mobilization can allow anastomosis above the inferior pulmonary vein. Free jejunal interpositions with microvascular anastomosis can be used for cervical reconstruction.

Pyloric drainage after gastric tube reconstruction

Meta-analysis of six comparative studies showed that pyloric drainage procedures showed a non-significant trend towards fewer anastomotic leaks, pulmonary complications and reduced gastric stasis when employed following oesophagectomy. However, the ideal technique remains unproven.

The practice of the authors is to test the emptying of the gastric conduit in the course of a postoperative contrast study that is carried out to test the integrity of the anastomosis. If delayed emptying through the pylorus is encountered, balloon dilatation under radiological guidance is carried out.

Multimodality oncological management

A meta-analysis of 24 studies including 4188 patients was reported in 2011 examining survival benefits of neoadjuvant chemotherapy and neoadjuvant chemoradiotherapy compared with surgery alone in oesophageal cancer. This meta-analysis provides strong evidence for a survival benefit of neoadjuvant chemoradiotherapy or chemotherapy over surgery alone in patients with oesophageal carcinoma. A clear advantage of neoadjuvant chemoradiotherapy over neoadjuvant chemotherapy has not been established.

Neoadjuvant chemotherapy

The UK Medical Research Council (MRC – EO2) randomized controlled trial of 802 patients into surgery alone compared with two 3 week cycles of cisplatin + 5-fluorouracil chemotherapy followed by surgery showed that the neoadjuvant chemotherapy group had better 2 year survival (43% vs 34%) and 5 year survival (23% vs 17%) than surgery alone.

This survival benefit was not observed in the US Intergroup-0113 study, which randomized 467 patients into surgery alone or three cycles of cisplatin + 5-fluorouracil chemotherapy followed by surgery and postoperative cisplatin + 5-fluorouracil chemotherapy for responders. However, there was a significant improvement in 5 year survival for patients who underwent R0 resection vs patients who had resection margins that were either microscopically or grossly positive (32% vs 5%). Also, patients who underwent incomplete resections (R1 or greater) had survival rates that were equivalent to those who did not undergo surgery at all. These findings highlight the importance of negative resection margins in combined modality treatment. Furthermore, patients who had an objective response to neoadjuvant chemotherapy had significantly better survival rates.

Several factors may account for the difference in results between those two trials using similar chemotherapeutic regimens. The MRC trial larger sample size might have allowed the detection of statistically small yet clinically significant survival advantage that could not be detected with a small sample size trial. Also, the histology for the two trials was different: about two-thirds of patients in the MRC trial had adenocarcinoma compared with about 54% in the Intergroup trial.

The perioperative chemotherapy MAGIC trial using three cycles of preoperative and postoperative epirubicin (E) 50 mg/m^2, cisplatin (C) 60 mg/m^2 and continuous intravenous infusion of 5-fluorouracil (F) 200 mg/m^2 per day (ECF) significantly improved 5-year survival from 23% with surgery alone to 36.3%. In this trial, 26% of patients had distal oesophageal or gastro-oesophageal adenocarcinoma. This regimen is currently the standard practice in the UK for those tumours.

Neoadjuvant chemoradiotherapy

A systematic review of 38 papers, comprising 3640 patients, published between 2000 and 2008 showed that neoadjuvant chemoradiation regimens for oesophageal cancer vary widely with a predominance of 5-fluorouracil/cisplatin chemotherapy. Chemoradiation-related toxicity was reported in only 10 studies and consisted mainly of neutropenia. The chemoradiation-related in-hospital mortality rate was 2.3% compared with 5.2% for oesophagectomy following chemoradiation. The mean R0 resection rate and pathological complete response rate were 88.4% and 25.8% respectively. Five-year survival rates varied from 16% to 59% in all patients and from 34% to 62% in those with a complete pathological response. Chemoradiation had a temporary negative effect on quality of life.

The question is whether neoadjuvant chemoradiotherapy increases the morbidity and mortality of oesophagectomy. Studies have demonstrated that oesophagectomy can be performed safely with low mortality and acceptable morbidity following neoadjuvant chemoradiotherapy with no increase in hospital stay for patients with complications.

A recent randomized trial (CROSS study) for resectable (T2-3N0-1M0) oesophageal tumours randomized 363 patients into preoperative chemoradiotherapy consisting of weekly administrations of paclitaxel and carboplatin for 5 weeks and concurrent radiotherapy (41.4 Gy in 23 fractions, 5 days per week) followed by surgery vs surgery alone. The complete pathological response was 32.6%. In-hospital mortality was 3.7% in the surgery alone arm vs 3.8% in the chemoradiotherapy arm. The toxicity of this regimen was also acceptable. The overall survival was significantly better in the group of patients treated with neoadjuvant chemoradiotherapy. Median survival was 49 months in the chemoradiotherapy arm vs 26 months in the surgery alone arm. One, 2 and 3 year survival rates are 82%, 67% and 59%, respectively, in the chemoradiotherapy arm and 70%, 52% and 48%, respectively, in the surgery alone arm.

Squamous cell cancer has a better response to chemoradiotherapy than adenocarcinoma. This has been shown in subgroup analysis of prospective clinical trials. Tumour response to neoadjuvant chemotherapy is an independent prognostic factor in patients with squamous cell carcinoma.

Definitive chemoradiotherapy

The Intergroup RTOG 85-01 study was the landmark study of definitive radiotherapy with or without chemotherapy in patients with resectable squamous or adenocarcinoma of the oesophagus (T1–3N0–1M0) not subjected to routine oesophagectomy. Patients were randomized to four cycles of 5-fluorouracil and cisplatin with concurrent 50 Gy total dose radiotherapy compared with 64 Gy total dose radiotherapy alone. Five-year survival was 26% in the chemoradiotherapy arm vs 0% in the radiotherapy alone group and the survival benefit was maintained at 8 year follow-up (22% vs 0%). Persistence of disease despite therapy was the most common mode of treatment failure, but it was less common in patients receiving combined therapy (26%) than in the group treated with radiotherapy alone (37%). Severe acute toxic effects were greater in the combined therapy groups but there were no significant differences in severe late toxic effects between the groups.

Two European randomized clinical trials have investigated the value of surgery in a combined modality approach to oesophageal cancer. All of the patients in the Stahl study and 90% of the patients in the Bedenne trial had squamous cell carcinoma. In the Stahl study, patients with locally advanced squamous cell carcinoma of the oesophagus were randomly allocated to either induction chemotherapy followed by chemoradiotherapy (40 Gy) followed by surgery, or the same induction chemotherapy followed by chemoradiotherapy (at least 65 Gy) without surgery. The study concluded that 'adding surgery to chemoradiotherapy improves local tumour control but does not increase survival of patients with locally advanced oesophageal SCC. Tumour response to induction chemotherapy identifies a favourable prognostic group within these high-risk patients, regardless of the treatment group'.

All patients in the Bedenne trial received two cycles of fluorouracil and cisplatin and either conventional (46 Gy in 4.5 weeks) or split-course (15 Gy, days 1–5 and 22–26) concomitant radiotherapy. Patients with response and no contraindication to either treatment were randomly assigned to surgery or continuation of chemoradiation [three cycles of fluorouracil/cisplatin and either conventional (20 Gy) or split-course (15 Gy) radiotherapy]. The 2-year local control rate was better in the surgical arm than in the definitive chemoradiotherapy group (66.4% vs 57.0%) and stents were also required less in the surgery arm (5% vs 32%).

From both trials, response to induction therapy should be the deciding factor as to who proceeds to surgery, as patients with a good response to therapy had equal (or better) survival to the group that eventually underwent surgery. Surgery provided better local control vs chemoradiotherapy alone, and, for those patients who did not respond to induction therapy, complete surgical resection improved survival. Nevertheless, oesophagectomy should be performed with low mortality rates, otherwise the survival benefits will be mitigated.

Palliative management of oesophageal cancer

First-line palliative chemotherapy

In patients with incurable disease, it is essential to consider patient comorbidity, functional status and quality of life. Patients with oesophageal cancer have the same benefits from palliative chemotherapy as those with gastric cancer and oesophagogastric junctional tumours. For patients who can withstand chemotherapy, the benefits of palliative chemotherapy over the best supportive measures have been shown in randomized controlled clinical trials.

In randomized controlled trials, combination therapy of ECF had superior response rates (45% vs 21%), median overall survival (8.9 vs 5.7 months) and 2 year survival (13.5% vs 5.4%) over FAMTX (5-fluorouracil, Adriamycin and methotrexate). Also, ECF had similar response and survival rates to MCF (mitomycin C, cisplatin and 5-fluorouracil) but better quality of life measures.

In the REAL-2 study, a two-by-two design, we randomly assigned 1002 patients to receive triplet therapy with epirubicin and cisplatin plus either fluorouracil (ECF) or capecitabine (ECX) or triplet therapy with epirubicin and oxaliplatin plus either fluorouracil (EOF) or capecitabine (EOX). The primary endpoint was non-inferiority in overall survival for the triplet therapies containing capecitabine compared with fluorouracil and for those containing oxaliplatin compared with cisplatin. Median survival times in the ECF, ECX, EOF and EOX groups were 9.9 months, 9.9 months, 9.3 months and 11.2 months, respectively; survival rates at 1 year were 37.7%, 40.8%, 40.4% and 46.8%, respectively. In the secondary analysis, overall survival was longer with EOX than with ECF, with a hazard ratio for death of 0.80 in the EOX group. Progression-free survival and response rates did not differ significantly among the regimens. The study concluded that capecitabine and oxaliplatin are as effective as fluorouracil and cisplatin, respectively, in patients with previously untreated oesophagogastric cancer.

In summary, patients with good performance status but inoperable oesophagogastric cancer should be offered combination therapy of EOX or ECX.

Target agent in combination with chemotherapy

In patients who have HER-2-positive oesophagogastric junctional or gastric tumours, the addition of trastuzumab to cisplatin and 5-fluorouracil resulted in significant improvement in response rate (47.3% vs 34.5%), progression-free survival (6.7 vs 5.5 months) and median overall survival (13.8 vs 11.1 months). Tumours are considered to be HER-2 positive if the immunohistochemistry score is 3+ or if the tumour is positive for fluorescent *in situ* hybridization.

Interventions for dysphagia

A Cochrane systematic review of 40 randomized controlled trials including 2542 patients with inoperable or unresectable primary oesophageal cancer who underwent palliative treatment was carried out. The review included different interventions such as rigid plastic intubation, self-expanding metallic stent insertion, brachytherapy, external beam radiotherapy, chemotherapy, oesophageal bypass surgery, and chemical and thermal ablation therapy. The primary outcome was improvement of dysphagia. Secondary outcomes included recurrent dysphagia, technical success, procedure-related mortality, 30-day mortality, adverse effects and quality of life (Figure 22.32).

The results of the analysis showed that self-expanding metallic stent insertion is safer and more effective than plastic tube insertion. Thermal and chemical ablative therapy provide comparable dysphagia palliation but have an increased requirement for reinterventions and adverse effects. Antireflux stents provide comparable dysphagia palliation to conventional metal stents. Some antireflux stents might reduce gastro-oesophageal reflux compared with conventional metal stents. Brachytherapy might be a suitable alternative to self-expanding metallic stents in providing a survival advantage and possibly a better quality of life.

The review concluded that ‘self-expanding metal stent insertion is safe, effective and quicker in palliating dysphagia than other modalities. However, high-dose intraluminal brachytherapy is a suitable alternative and might provide additional survival benefit with a better quality of life. Self-expanding metal stent insertion and brachytherapy provide comparable palliation to endoscopic ablative therapy but are preferable due to the reduced requirement for reinterventions. Rigid plastic tube insertion, dilatation alone or in combination with other modalities, chemotherapy alone, combination chemoradiotherapy and bypass surgery are not recommended for palliation of dysphagia due to a high incidence of delayed complications and recurrent dysphagia’.

Benign oesophageal tumours

Benign tumours of the oesophagus are part of a diverse and relatively rare group of pathologies which present infrequently as part of the general surgeon’s operative case mix. It is however important to recognize these tumours and distinguish them from life-threatening or malignant lesions, with which they may

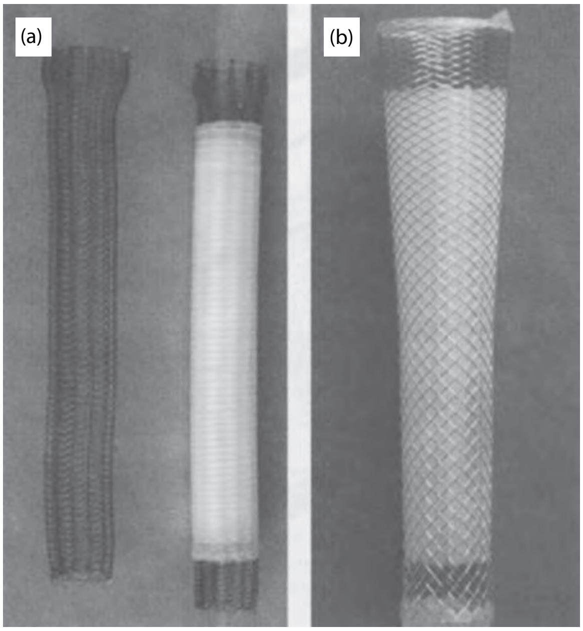


Figure 22.32 Palliation of dysphagia by stenting. (a) Ultraflex stents (covered and uncovered); (b) covered Flamingo stent.

coexist and intermittently be mistaken, and treat them with the aim of restoring oesophageal integrity and function, preferably through a minimally invasive approach.

Although several hundred cases of benign tumours have been reported over the past centuries under a plethora of names, their true prevalence is unknown. A number of autopsy reviews in the past century have reported a prevalence of 0.17–0.59% with the mid- to lower oesophagus being the most common sites of occurrence. More certainly, retrospective reviews have shown these tumours to constitute less than 10% of all resected oesophageal specimens.

Perhaps due to their paucity and range, classification of the benign tumours is also extremely varied. Historically, these disorders have been grouped according to the tissue histology (Box 22.7) or their relative location in the oesophageal wall (Box 22.8).

Clinical features of benign oesophageal tumours

Benign oesophageal tumours, although usually asymptomatic, are recognized to have a large range of presenting features (Box 22.9):

- *Asymptomatic.* The great majority of benign tumours are slow growing, arising in the submucosa or intramural portion of the oesophagus. As a result they tend to be insidious in nature, being discovered only incidentally during endoscopic or radiological investigation of the upper gastrointestinal tract. Owing to their nature, many patients with benign lesions may never be aware of their condition, which may account for the significant portion of cases in the premodern era that were found at autopsy without any obvious ill effect.
- *Obstruction from intraluminal growth.* Obstructive symptoms are usually the outcome of intraluminal lesion growth or extension from the oesophageal wall. Clinical symptoms tend to occur late in the

BOX 22.7 Classification of benign oesophageal tumours by cell type

- Epithelial
- Squamous cell papilloma
- Fibrovascular polyp
- Adenoma
- Inflammatory pseudotumour
- Inflammatory polyp
- Non-epithelial
- Leiomyoma
- Haemangioma
- Fibroma
- Neurofibroma
- Schwannoma
- Rhabdomyoma
- Lipoma
- Lymphangioma
- Hamartoma
- Heterotopic
- Granular cell tumour
- Chondroma
- Osteochondroma
- Giant cell tumour
- Amyloid tumour
- Eosinophilic granuloma

Adapted from Choong CK, Meyers BF. Benign esophageal tumors: introduction, incidence, classification and clinical features. *Semin Thorac Cardiovasc Surg* 2003;**15**:3–8.

BOX 22.8 Classification of benign oesophageal tumours by location

Intraluminal
- Fibrovascular polyp
- Squamous cell papilloma
- Inflammatory pseudopolyp
- Lipoma
- Fibroneuroid tumour

Mural
- Leiomyoma
- Inclusion cysts
- Rhabdomyoma
- Lipoma
- Hamartoma
- Haemangioma
- Granular cell tumour
- Neurofibroma

Extramural
- Duplication cyst

Adapted from Choong CK, Meyers BF. Benign esophageal tumors: introduction, incidence, classification and clinical features. *Semin Thorac Cardiovasc Surg* 2003;**15**:3–8.

BOX 22.9 Clinical presentations of benign oesophageal tumours

- Asymptomatic
- Obstruction from intraluminal growth
- Compression of adjacent tissue by extramural tumour
- Regurgitation of pedunculated tumour
- Ulceration and bleeding

disease process, usually when the tumours are large (>5 cm), resulting in varying levels of dysphagia, vomiting, cough and substernal discomfort. Although the most common feature in the symptomatic patient, dysphagia in association with benign oesophageal tumours occurs only sporadically. Moersch and Harrington, in a large retrospective review of 11 000 patients presenting with dysphagia from any cause, found benign tumours to be the underlying pathology in only 15 patients (0.14%).

- *Compression of adjacent structures by extramural tumour.* Benign tumours may occasionally grow to such a size that they result in compression of mediastinal structures and subsequent atelectasis, airway obstruction and superior vena cava syndrome. Furthermore, as benign tumours can coexist with other oesophageal disorders such as malignancy, achalasia, diverticulum and hiatal hernia, it is of critical importance to exclude any additional or alternative pathology to ensure the optimum operative outcome.
- *Regurgitation of pedunculated tumour.* Tumours arising from the smooth muscle of the muscularis mucosae can pedunculate and protrude into the oesophageal lumen. Pedunculated tumours are occasionally found following regurgitation into the upper airway where they can cause obstruction and death in severe cases.
- *Ulceration and bleeding.* Vascular lesions such as haemangiomas or pedunculated tumours with a large vascular pedicle may spontaneously ulcerate, resulting in acute bleeding. Additionally, endoscopic examination and histological sampling of a similar lesion will also produce acute haemorrhage.

Treatment of benign oesophageal tumours

Most authors advocate excision of the tumour through enucleation, regardless of the tumour type, due to the potential for coexisting malignant disease and risk of complication.

Regardless of the approach taken, a preoperative work-up aiming to characterize the lesion using a combination of endoscopy, barium swallow and CT scan is essential, thereby allowing exclusion of any coexisting comorbidity, malignant change and extraoesophageal involvement.

Treatment typically falls into one of four categories:

- observation
- endoscopic excision/ablation
- open surgery (laparotomy/thoracotomy)
- minimally invasive surgery (laparoscopy/thoracoscopy).

Observation

As most tumours are asymptomatic, slow growing and generally incidental findings, it has been advocated that, following exclusion of significant comorbidity and malignancy, it is occasionally possible to actively observe patients and intervene only if symptoms develop or if a further pathology requires operative treatment.

Endoscopic excision/ablation

As with the excision of colonic polyps, it has been advocated that the endoscopic approach would be feasible for removing oesophageal lesions. In particular, where tumours arise from the muscularis propria, the presence of a normal mucosal layer overlying the tumour permits its enucleation whilst preserving the submucosa. As with other polyps, a saline injection is used to lift the tumour away from the submucosa and delineate it more precisely, thereby allowing polypectomy via a snare approach. The procedure is associated with bleeding, which can usually be controlled through the use of an argon plasma beam or endoscopic clips. Conversion rates specifically for endoscopically removed benign oesophageal tumours are not known but the indications and incidence should typically be the same as for other endoscopic procedures.

Open surgery (laparotomy/thoracotomy)

Surgical approaches in the pre-endoscopic era typically involved either laparotomy or (right sided) thoracotomy. The precise approach depended on the location of the pathology. A thoracotomy was the preferred modality as it allowed the oesophagus to be accessed throughout its length from the cricopharyngeus to the diaphragmatic hiatus. Additionally, large tumours are more easily enucleated through the open approach and where necessary, in rare instances, oesophagectomy can also be performed.

Minimally invasive surgery (laparoscopy/thoracoscopy)

The minimally invasive approach has largely superseded the open approach over the past 20 years, particularly when operating on small or medium-sized lesions (typically up to 7–8 cm in length).

Patients are usually placed in the left lateral decubitus (for right-sided access) or prone position and intubated with a double-lumen endotracheal tube to permit single lung ventilation. An intraoperative, on-table endoscopy is advocated by some authors to identify the site of the lesion more accurately, determine the presence of any fibrosis surrounding the tumour which may complicate surgery by obliterating the anatomical planes, and ensure the mucosa remains intact postoperatively.

Once access is gained, the tumour is identified and myotomy performed, developing the plane between the tumour, muscularis propria and submucosa leading to eventual enucleation and excision. Some authors will endoscopically confirm that the mucosa is intact and, following this, the muscle layer is sequentially closed to complete the procedure.

In cases where a lesion lies in close proximity to the hiatus or GOJ, a laparoscopic approach is occasionally taken to excise the tumour and, where necessary, a fundoplication procedure is further conducted to prevent postoperative reflux.

The remainder of this section will focus on the common benign oesophageal tumours and their specific presentation and management.

Leiomyoma

Oesophageal leiomyomas are the commonest benign oesophageal tumours accounting for more than 50% of all such lesions at both autopsy and surgery. The oesophagus is the site of 10–12% of all gastrointestinal leiomyomas with a peak incidence between the third and sixth decades and a reported propensity for men more than women (1.8:1).

Pathology

Leiomyomas are usually singular (but multiple in 2.4% of cases), arising largely from smooth muscle cells and, due to the presence of striated (skeletal) muscle in the upper oesophagus, are found mainly in the mid- and lower oesophagus (Box 22.10). Tumours can very occasionally extend into the mediastinum, causing compressive symptoms. In a series of 838 cases, 97% of leiomyomas were found to be intramural, 1% polypoid intraluminal and 2% extraoesophageal.

Histologically they are composed of smooth muscle arranged in elongated, interconnected spindle cells with blunted nuclei and eosinophilic cytoplasm surrounded by hypovascular connective tissue. Although malignant change has been reported within leiomyomas, it appears to be a rare complication.

BOX 22.10 Anatomical location of oesophageal leiomyomas

- Upper oesophagus (10%)
- Mid-oesophagus (40%)
- Lower oesophagus (50%)

Clinical features

As is the case with benign oesophageal tumours in general, the great majority of leiomyomas are also asymptomatic, with symptoms typically occurring only when the tumour is >5 cm in size and causing a degree of luminal occlusion or mediastinal mass effect. The symptom range is extremely wide and non-specific (Box 22.11), with dysphagia and retrosternal discomfort the most common. Odynophagia results from oesophageal stretching or is due to the passage of a food bolus in the face of a compressed oesophageal lumen.

Diagnosis

As a result of the predominant intramural growth and limited intraluminal effect, diagnosis of leiomyoma is usually incidental during contrast studies, radiology or endoscopic examination of the upper gastrointestinal tract.

- *X-ray* can occasionally show a mediastinal mass particularly when calcified but is not in itself diagnostic.
- *Contrast studies* will show (1) a crescent-shaped filling defect within the oesophagus, (2) flattening of the mucosal folds overlying the tumour, (3) a mobile tumour mass on swallowing, (4) sharp demarcation of the tumour poles with the oesophageal wall and (5) splitting of the barium column as on passing the lesion (Figure 22.33).
- *Computed tomography* will further delineate the margins of the tumour and where malignancy or extraoesophageal involvement is suspected, would confirm the exact mass effects and nodal involvement (if any).
- *Upper gastrointestinal endoscopy* will most commonly show a round, moveable submucosal mass covered by a normal, easily displaced mucosa. Biopsy would confirm the lesion to be benign in nature.

BOX 22.11 Presenting features in symptomatic leiomyomas

- Dysphagia/odynophagia
- Retro/substernal pain
- Hypersalivation
- Postpranidal epigastric pain
- Dyspepsia
- Weight loss
- Bloating
- Cough
- Wheeze
- Dyspnoea
- Anaemia
- Regurgitation
- Nausea
- Vomiting

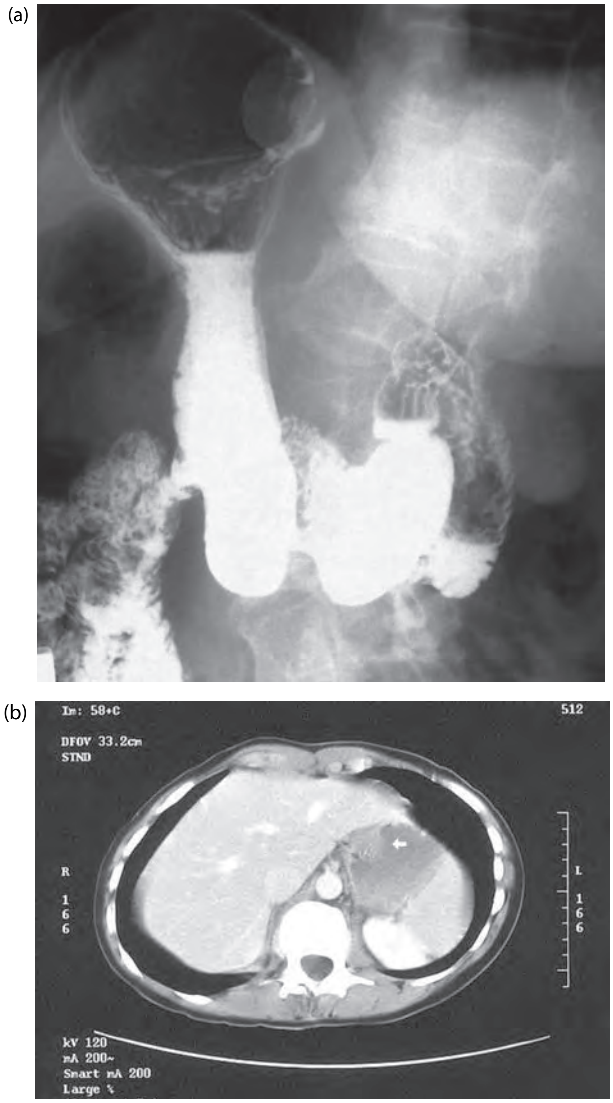


Figure 22.33 (a) Barium swallow showing a smooth ovoid tumour at the gastro-oesophageal junction. The lesion was a benign leiomyoma. (b) CT scan of the same patient.

Treatment

Surgical treatment consists of enucleation and excision according to the principles described previously. It is important to bear in mind that leiomyomas can coexist with diseases such as malignancy, diverticulum and hiatal hernia in which surgical treatment will usually be necessary.

Oesophageal inclusion and duplication cysts

Oesophageal cysts are the second most common benign oesophageal tumours, having an autopsy prevalence of one in 8200.

Inclusion cysts tend to be intramural and are often found in the upper oesophagus, in close proximity to the tracheal bifurcation. They consist of a mucosal lining of ciliated columnar or stratified squamous epithelium and contain a clear mucoid or brownish fluid. Oesophageal duplication cysts, where a portion of the oesophagus is doubled, are by comparison much rarer than inclusion cysts. Twenty per cent of gastrointestinal tract duplications occur in the thorax and some 60% of thoracic duplication cysts occur in the lower third of the oesophagus.

Clinical features

The effects of the cysts are usually the result of pressure exerted in the mediastinum. Particularly when in proximity to the tracheobronchial system, inclusion cysts can cause airway obstruction and recurrent respiratory tract infection.

Duplication cysts are extramural lesions, sharing a common muscular wall with the native oesophagus, extending for a variable distance parallel to the oesophagus with a predominantly gastric mucosa lining. The mucosa is associated with acid production which can subsequently result in ulceration, haemorrhage and oesophageal perforation.

Owing to the variety of potential complications that result from inclusion and duplication cysts, surgical excision and reformation of the oesophagus is the most widely practised treatment.

Fibrovascular polyp

These have been variously labelled in the medical literature as fibroma, fibrolipoma, myxofibroma, benign oesophageal polyp, fibroepithelial polyp and pedunculated lipoma. Fibrovascular polyps are the commonest intraluminal benign oesophageal tumours, mainly seen in men in their fifties and sixties.

Pathology

Although able to develop at any level, the normally solitary fibrovascular polyp is most often (>85%) found in the upper oesophagus lumen or wall, distal to the cricopharyngeus. The lesions usually start as small polypoid mucosal tumours which, with time, become elongated and pedunculated due to peristalsis and repeated swallowing. Seventy-five per cent of polyps are 7 cm or longer when patients present, with lengths of up to 20 cm being reported.

Microscopically, the polyps are thin, composed of loose fibrovascular tissue with mucosal folds or large elongated pedicles containing varying levels of blood vessels, adipose cells and stroma covered by normal squamous mucosa.

Clinical features

In the symptomatic patient, dysphagia is the usual presenting symptom. A review of 16 patients with symptomatic fibrovascular polyps found 87% and 25%, respectively, to have dysphagia and respiratory symptoms such as cough, wheeze and stridor. If sufficiently large, the distal tip of the polyp may ulcerate, resulting in bleeding and pain. Patients have infrequently been reported to present with tumour regurgitation into the mouth prior to the condition being recognized which can be associated with life-threatening airway obstruction.

Diagnosis

As with all benign tumours, diagnosis is largely coincidental. Most dramatically in the case of the fibrovascular polyp, it presents with oral regurgitation. More conventionally, the lesion, when combined with digestive or respiratory symptoms can be recognized on radiographs, which can show a relatively radiolucent oesophageal neoplasm. This may be more clearly delineated on contrast examination or CT. Endoscopy may also reveal the diagnosis, but if present intramurally the polyp may not always be evident.

Treatment

Fibrovascular polyps, once diagnosed, are treated with surgical resection to prevent possible oral regurgitation and fatal asphyxia due to laryngeal impaction. Small lesions can generally be excised endoscopically; however, the presence of a highly vascular pedicle can necessitate a cervical incision or even thoracic intervention. A minimal access approach is potentially possible for a polyp that is bound in the intramural space and can be freed up using a small myotomy, or with dissection within the oesophageal wall through a cervical incision.

Squamous cell papilloma

Squamous cell papillomas are small (typically measuring <1.5 cm), rare lesions (autopsy incidence of 0.01–0.04%) which occur usually in the elderly population.

Occasionally, multiple papillomas can occur, especially in the lower oesophagus, having the appearance of a mass of sessile warty lesions covered by a squamous mucosa with a fibrous core. This mass appearance of papillomas can be confused with a verrucous squamous cell carcinoma, with only biopsy providing the definitive diagnosis.

Histologically, the tumours show multiple finger-like projections of hyperplastic squamous epithelium with a central connective tissue core. The tumour has been associated with human papilloma virus and a single case of malignant degeneration has been reported, but overall this seems to be a rare complication. Treatment will be based on excision along established lines as discussed previously.

Granular cell tumour

These are rare lesions, approximately 0.5–2 cm in size, firm in consistency, arising from the submucosa and appearing as submucosal yellow, sessile polypoid structures with an intact mucosa, most frequently in the distal oesophagus.

The most common anatomical locations are:

- tongue (40%)
- skin (30%)
- breast (15%)
- alimentary tract (5%).

Oesophageal granular cell tumours account for 1–2% of all tumour cases.

Histologically, the tumours consist of polygonal cells with a distinct cellular border and pale cytoplasm filled with eosinophilic granules. Occasionally the tumours may be mistaken for squamous cell carcinoma because of pseudoepitheliomatous hyperplasia of the overlying mucosa.

Granular cell tumours are reported to stay stable over an extended period of time without much growth. However, malignant change has been reported in 3.4% of cases, suggesting that active resection is a preferable approach to that of watchful waiting.

Inflammatory pseudotumours

These are localized pedunculated tumours arising from the mucosa and found mainly in the distal oesophagus. They are associated with marked inflammatory changes within and may resemble a malignant carcinoma. Histological differentiation from a malignant process can be difficult but is vitally important to distinguish the two as pseudotumours do not usually require any specific treatment.

Inflammatory polyps

Inflammatory polyps result from reflux oesophagitis appearing macroscopically as oedematous inflamed gastric folds at the GOJ. Histologically, the polyp will demonstrate non-specific inflammatory infiltrates in the gastric mucosa, which contain plasma cells, eosinophils, multinucleated giant cells and fibroblasts. Managing the underlying gastro-oesophageal reflux will usually also be definitive for treating the polyp itself.

Adenomatous polyps

These lesions arise as a result of hyperplasia involving the epithelial cell itself, basement membrane and underlying vascular and connective tissues. They may additionally arise as a segment of columnar epithelium (Barrett's) with dysplastic change, which may occasionally be of a high grade requiring surgical excision.

Haemangioma

Haemangiomas result from hypertrophy of submucosal vascular tissue, appearing as solitary blue nodules, and account for some 2–3% of all benign tumours.

Patients are usually asymptomatic but can present with spontaneous bleeding if the overlying mucosa becomes ulcerated or following endoscopic biopsy or trauma.

Owing to the potential for unimpaired growth and haemorrhage, localized lesions are usually considered for surgical resection. However, one series examining the incidence of oesophageal haemangiomas reported 16 such lesions of which 12 were found at autopsy. This would suggest that surveillance is appropriate and treatment indicated if the lesions enlarge or bleed.

Summary

Benign solid tumours of the oesophagus are a rare but interesting selection of pathologies, which are usually diagnosed as incidental findings on endoscopic examination of the oesophagus. Although surgical treatment usually produces excellent results, it is vital that a thorough, accurate diagnostic work-up is conducted to exclude any coexisting pathologies and malignancy and ensure an optimum patient outcome without significant morbidity.

GUIDE TO FURTHER READING

Akiyama H, Tsurumaru M, Udagawa H, *et al.* Radical lymph node dissection for cancer of the thoracic esophagus. *Ann Surg* 1994;**220**:364–374.

Bedenne L, Michel P, Bouche O, *et al.* Chemoradiation followed by surgery compared with chemoradiation alone in squamous cancer of the esophagus: FFCD 9102. *J Clin Oncol* 2007;**25**:1160–1168.

Bello B, Herbella FA, Patti MG. Evolution of the minimally invasive treatment of esophageal achalasia. *World J Surg* 2011;**35**:1442–1446.

Boeckxstaens GE, Annese V, Bruley des Varannes S, *et al.* Pneumatic dilation versus laparoscopic Heller's myotomy for idiopathic achalasia. *N Engl J Med* 2011;**364**:1807–1816.

Bredenoord AJ. Mechanisms of reflux perception in gastroesophageal reflux disease: a review. *Am J Gastroenterol* 2012;**107**:8–15.

Broeders JA, Mauritz FA, Ahmed Ali U, *et al.* Systematic review and meta-analysis of laparoscopic Nissen (posterior total) versus Toupee (posterior partial) fundoplication for gastro-oesophageal reflux disease. *Br J Surg* 2010; **97**:1318–1330.

Choong CK, Meyers BF. Benign esophageal tumors: introduction, incidence, classification and clinical features. *Semin Thorac Cardiovasc Surg* 2003;**15**:3–8.

Cooper JS, Guo MD, Herskovic A, *et al.* Chemoradiotherapy of locally advanced esophageal cancer: Long-term follow-up of a prospective randomized trial (RTOG 85-01). Radiation Therapy Oncology Group. *JAMA* 1999;**281**:1623–1627.

Costantini M, Zaninotto G, Rizzetto C, *et al.* Oesophageal diverticula. *Best Prac Res Clin Gastroenterol.* 2004;**18**:3–17.

Demeester TR, Johnson LF, Joseph GJ, *et al.* Patterns of gastroesophageal reflux in health and disease. *Ann Surg* 1976;**184**:459–470.

Dent J, El-Serag HB, Wallander MA, Johansson S. Epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 2005;**54**:710–717.

DeVault KR, Castell DO. Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. *Am J Gastroenterol* 2005;**100**:190–200.

Eckardt AJ, Eckardt VF. Treatment and surveillance strategies in achalasia: an update. *Nat Rev Gastroenterol Hepatol* 2011;**8**:311–319.

Epstein D, Bojke L, Sculpher MJ. REFLUX trial group: Laparoscopic fundoplication compared with medical management for gastro-oesophageal reflux disease: cost effectiveness study. *BMJ* 2009;**339**:b2576.

Fennerty MB, Finke KB, Kushner PR. Short and long term management of heartburn and other acid-related disorders: development of an algorithm for primary care providers. *J Fam Pract* 2009;**58**:S1–S12.

Herskovic A, Martz K, al-Sarraf M, *et al.* Combined chemotherapy and radiotherapy compared with radiotherapy alone in patients with cancer of the esophagus. *N Engl J Med* 1992;**326**:1593–1598.

Jamieson GG, Lamb PJ, Thompson SK. The role of lymphadenectomy in esophageal cancer. *Ann Surg* 2009;**250**:206–209.

Johnson DA, Benjamin SB, Vakil NB, *et al.* Esomeprazole once daily for 6 months is effective therapy for maintaining healed erosive esophagitis and for controlling gastroesophageal reflux disease symptoms: a randomized, double-blind, placebo-controlled study of efficacy and safety. *Am J Gastroenterol* 2001;**96**:27–34.

Kamangar F, Chow WH, Abnet CC, Dawsey SM. Environmental causes of esophageal cancer. *Gastroenterol Clin North Am* 2009;**38**:27–57.

Kent M, d'Amato T, Nordman C, *et al.* Minimally invasive resection of benign esophageal tumors. *J Thorac Cardiovasc Surg* 2007:**134**;176–181.

Leyden JE, Moss AC, MacMathuna P. Endoscopic pneumatic dilation versus botulinum toxin injection in the management of primary achalasia. *Cochrane Database Syst Rev* 2006;**4**:CD005046.

Moersch HJ, Harrington SW. Benign tumors of the esophagus. *Ann Otol Rhinol Laryngol* 1944;**53**:800–817.

Nilsson M, Johnsen R, Ye W, *et al.* Lifestyle related risk factors in the aetiology of gastro-oesophageal reflux. *Gut* 2004;**53**:1730–1735.

Palanivelu C, Rangarajan M, Senthilkumar R, *et al.* Thoracoscopic management of benign tumors of the mid esophagus: a retrospective study. *Int J Surg* 2**007**:5;328–331.

Patti MG, Herbella FA. Achalasia and other esophageal motility disorders. *J Gastrointest Surg* 2011;**15**:703–707.

Richter JE. Oesophageal motility disorders. *Lancet* 2001;**358**:823–828.

Stahl M, Stuschke M, Lehmann N, *et al.* Chemoradiation with and without surgery in patients with locally advanced squamous cell carcinoma of the esophagus. *J Clin Oncol* 2005;**23**:2310–2317.

Sultan PK, Meyers BF, Patterson GA, *et al.* Fibrovascular polyps of the esophgaus. *J Thorac Cardiovasc Surg* 2005;**130**:1709–1710.

Symons NR, Purkayastha S, Dillemans B, *et al.* Laparoscopic revision of failed antireflux surgery: a systematic review. *Am J Surg* 2011;**202**:336–343.

Thomas ML, Anthony AA, Fosh BG, *et al.* Oesophageal diverticula. *Br J Surg* 2001;**88**:629–642.

Vakil N, van Zanten SV, Kahrilas P. The Montreal definition and classification of gastroesophageal reflux disease: a global evidence-based consensus. *Am J Gastroenterol* 2006;**101**:1900–1920.

Wang CY, Hsu HS, Wu YC, *et al.* Intramural lipoma of the esophagus. *J Chin Med Assoc* 2005;**68**:240–243.

Watson RR, O'Connor TM, Weisel W. Solid benign tumors of the esophagus. *Ann Thorac Surg* 1967;**4**:80–91.

Wen Z, Gardener W, Wang Y. Nitrates for achalasia. *Cochrane Database Syst Rev* 2004;**1**:CD002299.

Wileman SM, McCann S, Grant AM, *et al*. Medical versus surgical management for gastro-oesophageal reflux disease (GORD) in adults. *Cochrane Database Syst Rev* 2010;**17**:CD003243.

Wong TKH, Maydonovitch CL. Achalasia. In: Castell DO, Richter JE (eds) *The Esophagus*, 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins, 1999:185–213.

Yoshida N. Inflammation and oxidative stress in gastroesophageal reflux disease *J Clin Biochem Nutr* 2007;**40**:13–23.

CHAPTER 23

Disorders of the stomach and duodenum

SIR ALFRED CUSCHIERI AND GEORGE B. HANNA

Surgical anatomy

Macroscopic anatomy of the stomach

The stomach has an irregular pyriform shape tapering towards the duodenum and curved anteriorly so that its proximal (cardiac) orifice at the junction with the oesophagus and the distal end (pyloric sphincter) are at a more posterior plane near the retroperitoneum than the middle section (body) of the organ. It is anatomically divisible into three parts: the fundus, body and pyloric regions (antrum and pyloric canal). The fundus is the globular proximal portion to the left of the oesophagus separated from it by the cardiac notch and attached by fascia to the left crus and adjacent diaphragm and the gastrophrenic peritoneal fold. The pyloric region extends from the angular notch on the lesser curvature to the pyloric sphincter and consists of the antrum (proximal tapering portion) and the pyloric canal (tubular distal part just proximal to the pyloric sphincter), which is made up of a thickening of the circular muscle coat discontinuous with the equivalent muscle layer of the duodenum. The site of the pylorus is marked by a shallow superficial serosal notch and by the two veins of Mayo (superior and inferior) that cross its anterior surface. The anterior wall of the stomach is more voluminous that the posterior wall and this is important in antireflux surgery.

The normal stomach is lined by two types of mucosa:

- Parietal (oxyntic) lining the fundus and stomach contains parietal acid-secreting and chief cells that secrete pepsins. Surface mucosa has rugae and is acid.
- Antropyloric mucosa contains mucus-secreting and G-cells that secrete gastrin. Surface mucosa is smooth and neutral or slightly alkaline.

There are no surface markings for the antroparietal junction and the extent of the antral mucosa varies, but on average is approximately 5.0 cm from the pyloric orifice.

The stomach rests on the stomach bed consisting of retroperitoneal structures and organs: spleen, diaphragm, left kidney and adrenal, pancreas, splenic artery, transverse colon and mesocolon – hence the complexity of its vascular supply and lymph node drainage. It is almost completely lined by peritoneum anteriorly from the greater sac and posteriorly from the omental bursa (lesser sac). Above the lesser curvature of the stomach, these meet to form the lesser omentum, which is attached superiorly to the liver. The lesser omentum is usually thin; encloses the bile duct, hepatic artery and portal vein at its right extremity; and has a clear portion (devoid of fat) on the extreme left referred to as the pars flaccid that overlies the caudate lobe of the liver near the right crus of the diaphragm. The branches of the left and right gastric arteries run along the lesser curvature between the two layers of the lesser omentum. The two peritoneal layers meet along the greater curvature of the stomach to form a much larger and fattier fold, which has different names in its various sections, but which from the surgical standpoint should be considered as one structure sweeping from the entire margin of the greater curvature and fundus.

- Greater omentum – attached to the transverse colon and mesocolon, from which it can be separated through the embryonic vascular plane.

- Gastrosplenic ligament – contains the short gastric vessels to the spleen.
- Gastrophrenic ligament – small peritoneal fold that joins the stomach to the diaphragm to the left of the oesophagus.

Duodenum

Although anatomically part of the small intestine, the duodenum (the named is derived from its length being roughly equal to the breadth of 12 human fingers) is closely related to surgery of the stomach and pancreas. It is the shortest and widest part of the small intestine, has no mesentery and is only partially covered by peritoneum, its posterior aspect being fixed by fascia to the posterior abdominal wall. It forms an eccentric curve 28 cm long facing the left. Proximally it joins the stomach at the pylorus and distally the small intestine at the duodenojejunal junction, which is marked by a peritoneal fold known as the ligament of Treitz. It is divided into four portions: first (pars superior, first portion, duodenal bulb); second, descending portion or pars descendens varies from 7 to 10 cm in length and contains the duodenal papilla; third or horizontal portion (pars horizontalis, preaortic); and the fourth ascending portion (pars ascendens), which is short, averaging 2.5 cm in length. The duodenal bulb passes backwards from the pylorus before turning down to form the second part. The fourth parts bends forwards at the duodenojejunal junction. The superior relation to the duodenal curve is the gallbladder, the fundus of which touches the duodenal bulb. On the right side and posteriorly are the inferior vena cava, the right renal vessels, the renal pelvis/upper ureter of the right kidney and the aorta. Anteriorly, the second part is crossed by the transverse colon to which it is attached by loose areolar tissue, whereas the third part is crossed by the superior mesenteric vessels and the root of the small bowel mesentery. The head of the pancreas nestles inside the duodenal curve. In the groove between the duodenum and the pancreas runs the gastroduodenal artery, which arises from the common hepatic and descends behind the duodenal bulb to reach the groove.

The duodenal mucosa contains *Brunner's* glands, which secrete highly alkaline mucus that acts as a protective mechanism to the mucosa from the acid chyme delivered into the duodenum by the stomach. Usually the Brunner epithelium stops just beyond the duodenal papilla and never reaches the duodenojejunal junction.

Blood supply of the stomach and duodenum

Stomach

The stomach is supplied by the left gastric, right gastric, right gastroepiploic and left gastroepiploic arteries; thus, directly or indirectly, its blood supply is from the coeliac axis. The left gastric artery (origin from the coeliac axis or aorta) gives off an ascending oesophageal branch and then runs down between the two layers of the lesser omentum close to the lesser curvature to form an anastomotic arcade with the right gastric artery (origin from the hepatic artery). This arcade gives off delicate branches to the lesser curve but most of its bigger branches pierce the seromuscular coat anteriorly and posteriorly to join the submucosal vascular plexus of the stomach. Likewise, the right gastroepiploic (origin from the gastroduodenal artery) forms an arcade along the greater curvature and joins up with the left gastroepiploic (origin from the splenic artery). All the branches from the gastroepiploic arcade pierce the seromuscular layer to join the submucosal vascular plexus of the stomach with no direct branches to the greater curvature.

Extensive arterioarterial and arteriovenous connections are present within the gastric submucosal vascular plexus. The nutritional blood supply to the stomach (with the exception of the lesser curvature) is derived from this plexus. This accounts for the viability of the stomach when only one artery (right gastroepiploic) is retained during gastric tube reconstruction after oesophagectomy. It also underlies the increased safety of tubulization of the stomach, which entails removal of the lesser curve and adjacent stomach over reconstruction with the use of the intact stomach. The other benefit of tubulization is increased length. The venous drainage of the stomach parallels roughly the arterial supply.

Duodenum

The duodenal blood supply is through the duodenopancreatic arcades, which supply the duodenum and pancreas and form an important collateral route between the coeliac and superior mesenteric territories. The duodenopancreatic arcades are formed by the gastroduodenal artery (origin hepatic artery) and the anterior and posterior inferior pancreatic arteries (originating from the superior mesenteric artery). The gastroduodenal artery soon after its origin behind the duodenal bulb gives off the posterosuperior pancreaticoduodenal (or retroduodenal) artery, which curves round the lower common bile duct in its descent to the posterior aspect of the head of the pancreas and the duodenum. The gastroduodenal artery then continues in the groove between the duodenum and the pancreas and divides into the right gastroepiploic and anterior superior pancreaticoduodenal artery. The latter together with the retroduodenal artery anastomoses with the anterior and posterior inferior pancreaticoduodenal arteries derived from the superior mesenteric artery. This explains the difficulty in devascularizing the duodenum in bleeding duodenal ulcers. The collateral circulation between the coeliac and superior mesenteric arteries through the duodenopancreatic arcades also accounts for the unpredictable results of ligation of the common hepatic artery.

Nerve supply

The sympathetic supply is through the splanchnic nerves (T5/6–T9/10) with the preganglionic efferent fibres ending in the coeliac plexus. The postganglionic fibres then travel in the periarterial plexuses along the arteries to the stomach and duodenum. The sympathetic fibres subserve visceral sensation and pain. The parasympathetic supply is from the vagal nerves. There is considerable variation in the anatomical distribution of the vagal branches to the stomach and duodenum, which is not helped by the confusing eponymous nomenclature and lack of precise knowledge regarding the origin of some of the branches. Classically, the anterior vagus has the following branches:

- Hepatic branches – constant. They travel up towards the liver in the pars flaccid of the lesser omentum

- The pyloric nerve of McCrea – inconstant nerve to the pyloric region.
- Anterior gastric division or nerve of Latarget – constant antral branches which are preserved in highly selective vagotomy.

The posterior vagal trunk gives off the following branches:

- The posterior gastric division – posterior nerve of Latarget – smaller but essentially a mirror image of the corresponding anterior nerve
- The coeliac division to the coeliac plexus.

The proximal branches to the fundus have a variable origin from the nerves of Latarget, from the main vagal trunks at or above the hiatus and even from the oesophageal nerve plexus in the mediastinum. One of these nerves arising from the anterior trunk close to the oesophagogastric junction is known as the nerve of Grassi and is held responsible for some of the recurrences after vagotomy, although other proximal gastric branches are equally important is this respect.

In addition, there are gastroepiploic nerves which travel along the related vessels in the greater omentum and are said to innervate the parietal cell mass alongside the greater curve. There is no doubt regarding the existence of these nerves as they can be demonstrated in thin patients, but their origin and function has not been established, although some regard them as important in ulcer recurrence after highly selective vagotomy.

Most of the nerve supply to the duodenum is from the coeliac plexus.

Ultrasound of gastric mucosal layers and anatomy of the gastric mucosal glands

Endoscopic ultrasound (EUS) has enabled the acquisition of *in vivo* data from human subjects on the wall thicknesses of the stomach and other sections of the alimentary tract (needed in radiation dosimetry). Five layers of differing echo patterns are outlined EUS by of the stomach:

- superficial mucosa
- deep mucosa
- submucosa
- muscularis externa
- serosa/adventitia.

In a study of human patients, the thicknesses for stomach mucosa ranged from 1030 ± 130 to 1640 ± 80 m. The total stomach wall thicknesses varied from 2.80 ± 0.12 to 4.23 ± 0.03 mm.

The gastric mucosa, which is heaped up in rugae in the fasting state, is dotted with deep pits into which open the gastric glands, which vary in their morphology in accordance with their location in the stomach: cardiac glands at the cardia, fundic glands at the fundus and antropyloric glands. Different cell types are found at different levels (Figure 23.1) from the opening of the gland into the gastric pits: isthmus, body and base of gland.

- Mucous glands are located at the isthmus (neck) of the cardiac, fundic and antropyloric glands – secrete alkaline clear mucus which lines the surface of the stomach as a protective layer.
- Parietal (oxyntic) cells located in the body of the glands (mainly antropyloric) – secrete hydrochloric acid (HCl) and intrinsic factor (IF).

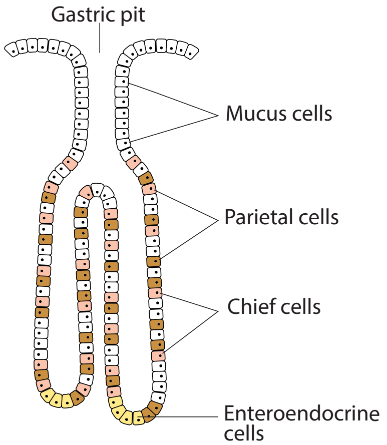


Figure 23.1 Schematic to illustrate the relation between the gastric pits and the gastric glands (cardia, fundic, antropyloric).

- Chief (peptic, zymogenic) cells located at the base of fundic glands only – secrete pepsinogens.
- Ghrelin endocrine cells located at the base of the fundic glands only – secrete ghrelin which stimulates release of growth hormone (see section Gastric physiology).
- Enteroendocrine cells also known as amine precursor uptake decarboxylase located at the base of all glands – secrete gastrin, histamine, endorphins, cholecystokinin (CCK), somatostatin.

Gastrin is released by the G-cells in response to distension of the antrum by ingested digested food (chyme). It stimulates the secretion of HCl from the parietal cells, and pepsinogen from the chief cells also increases gastric motility. Gastrin secretion is inhibited by a low pH (<4.0) and by somatostatin.

CCK induces contraction and thus emptying of the gallbladder after meals (especially fatty), decreases gastric emptying and induces the postprandial release of pancreatic juice which is alkaline and thus neutralizes the acid chyme entering the duodenum from the stomach.

Lymph node drainage

The drainage of lymph from the stomach is very complex and arranged in tiers and lymph node stations in the supracolic compartment and defined by the Japanese Society for Research in Gastric Cancer into 20 numbered stations which form the basis of modern curative resections for gastric cancer. The details of the lymph node stations and the various types of radical resections are outlined later on in this chapter (see Gastric cancer).

Gastric physiology

The most important physiological function of the stomach is to act as a receptacle (capacity 0.5–1.0 L) for ingested food, which is partially digested and hydrolysed by acid and pepsin, milled into semifluid chyme in the antropyloric region and then delivered in regulated aliquots in the postprandial period (gastric emptying). To enable the ingestion of an adequate amount of food, the stomach (particularly the fundus) possesses the property of adaptive relaxation or accommodation, which enables it to enlarge its capacity without a significant rise in the intraluminal pressure. The adaptive relaxation is mediated by stretch reflexes and vagal afferents and is therefore greatly impaired after vagotomy. It is also reduced to varying extents after partial gastrectomy and fundoplication. Loss of adaptive relaxation is manifested clinically as early satiety and can be caused by disease (usually malignant) that infiltrates the wall of the stomach The stomach is able to absorb most acidic drugs by passive diffusion of their lipid undissociated form, e.g. aspirin, other non-steroidal anti-inflammatory drugs (NSAIDs), ethanol, thiopental, secobarbital and antipyrine, whereas all substances which are almost completely ionized in acid solution are not. Ironically, some of the readily absorbed substances (NSAIDs, ethanol) are also known to cause gastric irritation and their overuse is commonly associated with development of gastritis and gastric ulcers.

The mucosa lining the body and fundus (proximal stomach) contains parietal (oxyntic), chief (zymogen secreting) and ghrelin-secreting G-cells (fundus only), whereas the mucosa lining the more muscular antropyloric segment secretes predominantly an alkaline mucus but contains specialized endocrine (G) cells that release gastrin in a co-ordinated fashion under the influence of gastrin-releasing peptide (GRP). The parietal cells secrete hydrochloric acid (HCl) and IF (necessary for the absorption of vitamin B_{12}) in the terminal ileum, whereas the chief cells are responsible for the production and secretion of pepsinogens that are then converted to active pepsins by HCl.

The motility of the stomach is a highly organized event controlled by the enteric nervous system. The initial adaptive relaxation after eating is followed by an increased activity of the antropyloric segment (antral mill), the contractions of which together with the acid/pepsin digestion convert the food into semifluid chyme before onward controlled passage into the duodenum via the pyloric sphincter. Normal gastric secretion consists of three interconnected phases: *cephalic, gastric* and *intestinal*. The *cephalic phase* follows stimulation of the vagal centres in the hypothalamus from the psychosensory input: expectation, sight, smell and chewing of food. The effect is mediated by direct cholinergic stimulation of the parietal cell mass and by cholinergic potentiation of other stimuli including histamine release. The *gastric phase* has two components. The first is consequent on intragastric stimulation by stretch receptors and chemoreceptors in the body that activate local and vasovagal reflexes evoking acid secretion. The second is the release of gastrin from the antrum by stimulation of stretch receptors and activation of antral chemoreceptors by peptides and amino acids released from the food acting via GRP. The *intestinal phase* is initiated by the entry of chyme into the duodenum when bile, pancreatic juice and various hormones are released. A delicate feedback inhibition occurs whereby acid and fat in the duodenum inhibit further gastric secretion by the release of secretin and CCK.

Peripheral mechanisms of gastric acid secretion

The secretion of acid is an important function of the human stomach, as by killing bacteria it provides a first-line defence against enteric infection. Gastric acid is also essential for the conversion of inactive pepsinogens to pepsins which initiate the digestion of dietary protein. Gastric acid is secreted by the human stomach constantly but the secretion rate varies, being low in the fasting state but sufficient to maintain the intragastric pH below 2.0. The secretion rate is highest with eating when the smell and taste of food stimulate the parietal cells mass via the vagal nerves. Additionally, the protein of ingested food stimulates the G-cells situated at the antral region to release gastrin, which circulates and stimulates the parietal cells in the body to secrete HCl. There are protective mechanisms which prevent excessive acidity of the stomach and duodenum. These include the release of somatostatin by D-cells within the antral mucosa, which exerts a paracrine inhibition of gastrin by the G-cells.

Gastric acid secretion is stimulated by acetylcholine, gastrin and histamine. Special receptors for these secretagogues are present on the surface of the gastric mucosal cells. The H_2 receptor (for histamine) is the most efficacious receptor in stimulating gastric secretion and has been cloned and shown to belong to the same family as the -adrenergic receptors. Its blockade by H_2-receptor antagonists (cimetidine, ranitidine, etc.) used to form the basis for acid suppression in the treatment of duodenal ulcer disease and reflux oesophagitis but has largely been replaced in clinical practice by proton pump inhibitors (PPIs).

Acetylcholine is released from the postganglionic neurones of the vagal sympathetic system and this is referred to as neurocrine secretion. Gastrin is released from the specialized G-cells into the bloodstream and exerts a hormonal action. Histamine is released from three sites (mast cells, enterochromaffin-like cells and nerves) into the interstitial fluid of the gastric mucosa and thus acts by a paracrine action.

The main inhibitor of gastric secretion is somatostatin, which is released from the D-cells and exerts a continuous inhibitory paracrine effect on the parietal cells of the fundus and on the G-cells of the antrum. Various hormones, such as secretin, CCK, vasoactive intestinal polypeptide (VIP), etc., inhibit gastric secretion through the release of somatostatin. More recent studies have shown that increased vagal cholinergic activity results in both direct stimulation of acid secretion and indirect stimulation by suppression of somatostatin restraint. This process is referred to as disinhibition. Other substances that inhibit gastric secretion include neurotensin, substance P and high concentrations of alcohol.

The acid pump

Irrespective of the chemical messenger (histamine, acetylcholine, gastrin) the ability of the parietal cells to secrete acid is dependent on the integrity of the enzyme H^+, K^+-adenosine triphosphatase (ATPase), which is also known as the acid pump. The system exchanges the H^+ from the intracellular water for extracellular K^+ present in the lumen of the canaliculus. Reversible inhibition of the ATPase pump can be achieved by PPIs, e.g. omeprazole. PPIs are inactive when administered but in the highly acid environment of the acid secretory canaliculus of the parietal cells, omeprazole is converted to cationic sulphenamide, which reacts covalently with groups in the outside of the pump producing long-lasting total but reversible inhibition of acid secretion. PPIs have largely replaced H_2 blockers in the treatment of peptic ulcer disease and reflux oesophagitis. However, with prolonged administration PPIs may cause a significant hypergastrinaemia. In experimental animals prolonged administration of PPIs results in the development of benign tumours of the enterochromaffin cells. The significance of this to prolonged therapy in the human remains uncertain although in practice only short or intermittent (on demand) PPI therapy is used.

Gastric secretion of pepsinogens

Pepsins are aspartate proteinases and are synthesized and secreted by the chief cells as prozymogens. The three important pepsins are pepsin I, pepsin II (gastricsin) and pepsin III (chymosin). Their main physiological role is acid-dependent digestion of dietary proteins within the stomach. The released peptides and amino acids mediate gastrin release, and through this mechanism the food-dependent gastric secretion of both acid and pepsins. Although pepsins may be involved in the pathophysiology of gastritis, peptic ulceration and gastro-oesophageal reflux disease, the exact mechanisms are still poorly defined, although digestion of the collagen IV component of the basement membrane, with delayed healing of any mucosal injury, may be operative.

Physiology of the gastric fundus

The primary function of the gastric fundus is to receive and store food. In the basal (fasting) state, the fundus is contracted, its resting tone being determined by a balance between excitatory cholinergic and inhibitory nitrergic drive. In humans activation of 5-HT_1receptors (by sumatriptan) relaxes the fundus. This relaxation is reversibly blocked by inhibition of nitric oxide synthase, confirming the nitrergic pathway. Variations in the tone of the gastric fundus alter the reservoir volume of the stomach during eating, thus enabling receptive relaxation. Additionally, gastric emptying is influenced by changes in the tone of the gastric fundus. Receptive relaxation during eating is largely mediated through activation of a nitrergic pathway. Efferent vagal preganglionic fibres synapse on both cholinergic excitatory and non-adrenergic non-cholinergic (NANC) inhibitory intrinsic neurones within the myenteric plexus of the fundus wall. The precise nature of the specific neurotransmitter released by the NANC neurones responsible for gastric relaxation is still uncertain, although on the basis of several experimental studies is thought to be nitrous oxide or VIP since both induce NANC-mediated relaxation of the fundus.

Growth hormone and ghrelin

Growth hormone (GH) has a complex regulation with two antagonistic hypothalamic hormones: growth hormone-releasing hormone and somatostatin, as well as the liver-derived hormone insulin-like growth factor (IGF)-I. Its secretion is influenced by metabolic signals (plasma levels of glucose, amino acids, free fatty acids, keto-acids) and the energy balance/nutritional state. Physiologically, GH has a dual action: somatic growth and the regulation of general metabolism.

The discovery of ghrelin followed the synthesis of artificial analogues referred to as growth hormone secretagogues (GHSs), including the highly potent GH-releasing hexapeptide, which stimulate the release of GH through their action on specific receptors in the pituitary gland. Following the cloning of the GHS receptor, this ligand was used to screen different tissue extracts. The highest expression of GHS receptor-activating factor, now known as the hormone ghrelin, is found in the gastric fundus where the ghrelin cells are located close to the capillaries and have no contact with the lumen of the oxyntic glands, indicating their endocrine nature. In most animals and in humans, the circulating ghrelin (normal plasma level range, 200–600 ng/L) originates largely from gastric fundal secretion. However, the majority of plasma ghrelin is not biologically active, as it is deamidated. Current radioimmunoassays measure total ghrelin. Through its release of GH, ghrelin is actively involved in the control of energy balance and the regulation of food intake.

The regulation of ghrelin secretion is controlled by food intake, vagal activity and various hormones. Somatostatin, cortistatin, thyroid hormones and insulin reduce gastric ghrelin secretion, whereas CCK and gastrin stimulate secretion (Figure 23.2). Human ghrelin is known to cross the blood–brain barrier and ghrelin administration activates fos and Egr-1 proteins in neurones of the arcuate, paraventricular and dorsomedial nuclei, and the area postrema of the hypothalamus. There is however controversy as to whether ghrelin acts by direct activation of central nervous system receptors or indirectly through activation of vagal nerve fibres. The latter mechanism has been suggested by the results of several animal studies, in which vagotomy was demonstrated to abolish ghrelin-induced feeding and GH secretion.

In humans, ghrelin administration (by either intravenous or direct infusion via the intracerebroventricular route) releases GH. Hence it is likely that gastric ghrelin may be involved in the physiological regulation of GH regulation secretion. Ghrelin administration in humans induces a sensation of hunger, an action which is opposite to that of the hormone leptin (derived from white adipose tissue). Ghrelin levels

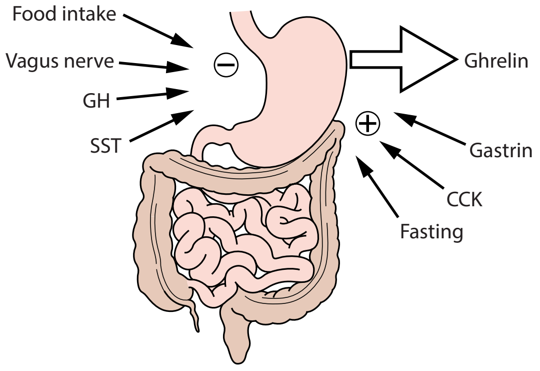


Figure 23.2 Regulation of gastric-derived ghrelin. GH, growth hormone; SST, somatostatin; CCK, cholecystokinin.

are decreased in obese patients and elevated in malnutrition states from any cause. Weight gain in malnourished patients is accompanied by a gradual reduction in the plasma ghrelin levels to normal. The plasma ghrelin levels are decreased after feeding in a reciprocal pattern with insulin, and in the fasting state, ghrelin levels are in phase with leptin. It seems therefore that the preprandial ghrelin rise may have a role in initiating meal consumption in humans. Obese patients who lose weight exhibit an increase in plasma ghrelin. This has been suggested as the reason for failure of low-calorie diets in the treatment of morbid obesity. In contrast, patients who have undergone bariatric surgery for morbid obesity have reduced ghrelin levels, thought to result from absence of direct food stimulation on the gastric fundus. It also probably explains why gastric bypass surgery is more effective in the long term than other bariatric operations. Additionally patients treated with gastric bypass usually experience reduced appetite after the operation. Physiologically, it is now thought that ghrelin anticipates the initiation of meals and releases GH. In malnutrition/catabolic states, the raised ghrelin levels through increased secretion of GH induces increased appetite and food intake, increased gastric emptying and food assimilation necessary for increased nutrient assimilation for restoration of muscle bulk and fat reserves. These actions are opposite to those of leptin, which reduces food intake and fat reserves.

Gastroduodenal mucosal defence mechanisms

The important defence mechanisms that protect the gastroduodenal mucosa against a variety of insults including drugs such as ethanol and NSAIDs are:

- surface mucous gel
- bicarbonate secretion
- reactive hyperaemia
- epidermal growth factors (EGFs)
- polyamines and ornithine carboxylase.

The outer surface of the gastric mucosa is covered by a mucous gel, which is hydrophobic. This property is dependent on phospholipid-containing vesicles, myelinated structures and a phospholipid band in the deeper layers of the mucous gel covering the luminal surface of the mucosa. The gastric mucosal cells extract bicarbonate from the mucosal blood and secrete it into the gastric lumen across the apical plasma membrane. Bicarbonate also reaches the gastric lumen by tracking between the gastric cells (paracellularly). This constant bicarbonate flux is reduced by smoking, certain drugs (e.g. cyclo-oxygenase inhibitors) and somatostatin. Enhanced secretion of bicarbonate is induced by prostaglandins and theophylline. When acid or other noxious agents enter the gastric mucosal cells, sensory neurones are stimulated, which then cause the release of vasodilatory neuropeptides (calcitonin gene-related peptides), nitric acid and prostaglandins with a resultant hyperaemia. This increased blood flow permits the neutralization of the back-diffusing acid before significant damage occurs. If the surface epithelium is damaged, the increased bicarbonate load brought to the mucosa by the reactive hyperaemia spills out and is trapped within the surface mucus, cellular debris and fibrin, forming a mucoid cap that covers the injured area, thereby providing a protective microenvironment with a stable pH 5–6, despite the high luminal acidity of the stomach.

EGF, which is produced by the salivary glands and Brunner's glands under physiological conditions, is concerned both with mucosal defence against injury and in the repair of established mucosal damage. Deficient secretion of EGF in patients with rheumatoid arthritis and Sicca syndrome is thought to be responsible for the enhanced susceptibility to NSAID-induced erosive disease exhibited by these patients. When ulceration develops anywhere in the gastrointestinal tract, special cells derived from the crypts adjacent to the ulceration site appear and secrete abundant amounts of EGF. This together with the synthesis of polyamines (required for cell growth and differentiation) by the enzyme *ornithine carboxylase* and other growth peptides is responsible for the reparative process.

Changes in the gastric mucosa induced by infection with *Helicobacter pylori*

In view of the importance of *Helicobacter pylori* infection in the aetiology of various benign and malignant disorders of the stomach, the hormonal, acid secretory changes leading to various morphological changes of the gastric mucosa induced by persistent infection by this organism are of importance in understanding the pathogenesis of these *H. pylori*-related disorders. It is now believed that *H. pylori* infection is usually contracted in childhood, causing acute abdominal pain due to severe inflammation of the antrum and body of the stomach associated with hypochlorhydria due to inhibition of the parietal cells. The infection then enters a chronic stage which persists throughout life, unless discovered because of a specific related disorder and eradicated by appropriate treatment. Undiagnosed, the chronic infection leads to morphological changes in the gastric mucosa. The effects on gastric secretion are determined by the extent to which the chronic gastritis affects the mucosa of the antrum or body of the stomach. Increased acid secretion in subjects with antral predominant disease is mainly due to *H. pylori* gastritis stimulating increased release of the hormone gastrin, which stimulates the parietal cells of the body of the stomach to secrete acid without the normal feedback restraint by antral somatostatin as these D-cells are destroyed or depleted by the antral gastritis. In subjects with a large parietal cell mass this leads to the development of duodenal ulceration.

When the chronic infection is predominantly located in the body of the stomach, hypochlorhydria ensues, with an increased risk of enteric infections due to the loss of the gastric acid barrier. Patients with chronic *H. pylori* body gastritis and hypochlorhydria or achlorhydria have in addition an increased risk of gastric cancer especially if the gastritis progresses to mucosal atrophy and intestinal metaplasia. The exact mechanism by which the hypochlorhydria associated with *H. pylori* body gastritis leads to the greatly increased risk of gastric cancer remains unclear, although one hypothesis concerns the production of potentially carcinogenic nitrosoamines following bacterial colonization of the stomach when the intragastric pH rises above pH 4.0, as various bacteria can produce carcinogenic nitrosamines from dietary nitrite and secondary amines.

In sharp contrast, the hypochlorhydria associated with chronic *H. pylori* body gastritis is now considered to have a beneficial effect in protecting subjects from acid-related upper gastrointestinal disorders, especially gastro-oesophageal reflux disease. This is based on the documented observation that the fall in prevalence of *H. pylori* in Western countries has been accompanied by a rise in prevalence of gastro-oesophageal reflux disease and increased incidence of Barrett's columnar cell metaplasia, dysplasia and junctional gastro-oesophageal cancer.

The important changes induced by chronic *H. pylori* infection in the gastric mucosa are summarized as follows:

- cell proliferation and apoptosis
- duodenal ulcer – antral-predominant disease
- increase in mucosa-associated lymphoid tissue (MALT) and MALT lymphomas
- gastritis, intestinal metaplasia, atrophy, gastric carcinoma – body-predominant disease.

Several clinical studies have established that eradication of chronic *H. pylori* infection by the appropriate therapy is followed within a period of 5 years by regression of intestinal metaplasia in the antrum and certainly no progression of the disease. Eradication therapy is therefore regarded as an effective prophylactic measure in patients at risk. In a seminal study from Japan involving patients with early gastric cancer, eradication following endoscopic resection significantly reduced the incidence of second cancers.

Physiological consequence of gastrectomy and gastric bypass

It is important to distinguish between the *physiological consequences* of gastric surgery (gastrectomy and gastric bypass) which occur in all patients and require supplements for their prevention and the adverse sequelae which occur in some but not all patients with varying severity and are collectively known as *postgastric surgery syndromes*. These often result in significant impairment of the quality of life of the affected patients and invariably pose problems in management. They are considered later in the chapter.

By and large gastric bypass and gastrectomy (total/subtotal) share the same physiological consequences although there are minor differences between the two. Loss of weight is invariable after total/subtotal gastrectomy and occurs in all patients after gastric bypass provided these patients adhere to the prescribed diet and undertake regular exercise. Weight loss after gastrectomy is marked with the development of malnutrition in patients who experience significant postcibal symptoms. In these patients, the resulting diminished dietary intake is the major factor and far outweighs others, such as malabsorption and decreased transit times. Mild steatorrhoea occurs in 70% of patients after gastrectomy, but severe steatorrhoea is rare (as distinct from gastric bypass patients) and is usually encountered in patients who develop bacterial overgrowth. Anaemia is a frequent adverse consequence of both gastrectomy and vagotomy with drainage, and its incidence increases with the duration of follow-up. Iron-deficiency anaemia accounts for the majority with an incidence of 60% in males and 75% in females at 10–20 years after gastric surgery. The higher incidence in females is due to the regular monthly menstrual losses. The exact pathogenesis of iron-deficiency anaemia remains unclear. Malabsorption of iron is likely to play a role, but several other factors have been incriminated: shift to trivalent ferric iron at high pH, loss of a gastric juice factor facilitating iron absorption, increased binding of dietary iron to proteins, diminished hydrolysis of iron–protein complexes in food and chronic blood loss from gastritis and erosions. Whatever the exact cause, sufficient oral iron can be absorbed even after total gastrectomy to restore the serum iron level to normal, and if prophylactic iron supplementation is administered (300 mg q.i.d.), the development of postgastrectomy iron-deficiency anaemia is prevented.

Vitamin B_{12} malabsorption is invariable after total gastrectomy, and unless replacement therapy is initiated megaloblastic anaemia develops within 3–4 years when the body stores of the vitamin become depleted. Although malabsorption of vitamin B_{12} is well documented after partial gastrectomy and gastric bypass surgery for morbid obesity, frank megaloblastic anaemia is rare. The main factor responsible in these patients is the lack of a sufficiently acid environment that is normally required to release vitamin B_{12} bound to food. Although IF secretion is reduced after partial gastrectomy, vagotomy and drainage and gastric bypass surgery, the residual secretion appears to be adequate in physiological terms in the majority but not all patients.

Postgastrectomy bone disease is less frequently encountered nowadays. Although osteomalacia accounts for the majority, cases with features of both osteoporosis and osteomalacia are well documented. The aetiology is multifactorial. Diminished dietary intake of calcium and vitamin D is important as is malabsorption. Postgastrectomy bone disease is usually encountered in elderly patients many years after total, subtotal and polygastrectomy as there is a latent period of many years. There is also an established female preponderance. The biochemical features (raised alkaline phosphatase and serum calcium) and radiological changes (refraction) predate the onset of symptoms, which include generalized bone pain, stress fractures and weakness from the associated myopathy.

Clinical features

Dyspepsia

The function of the stomach and duodenum is concerned with the initiation of the process of digestion. This is achieved by a combination of mechanical fragmentation and acid/peptic digestion of foodstuffs with an orderly delivery of the resulting acid chyme into the duodenum, where further chemical digestion in an alkaline medium continues. In health we are unaware of these activities, which are the result of co-ordinated secretory and motor functions by neural and hormonal mechanisms. Gastroduodenal disease disturbs many of these physiological mechanisms and produces varied symptoms described clinically under the embracing term of *dyspepsia*. Dyspeptic symptoms are however extremely common in the general population with reported prevalence rates varying from 14% to 44%, but show marked regional differences. It has been estimated that only 25% of patients with dyspepsia seek medical attention and in these flexible endoscopy is normal in a substantial but varying extent: 25–76%. One of the many reasons for this problem has been defining what constitutes dyspepsia in the general population. Several international working parties and consensus conferences have addressed this issue with some measure of, but not total, agreement on the constituent symptoms, severity grades and the appropriateness of endoscopy.

An agreed international definition of dyspepsia is 'episodic or persistent abdominal symptoms, often related to the intake of food, which patients or physicians believe to be due to disorders of the proximal portion of the digestive tract'. The symptoms included in this generic definition of dyspepsia agreed at the Maastricht Consensus Conference (1997) are:

- pain or discomfort in the upper abdomen
- nausea and vomiting
- early satiety
- epigastric fullness and regurgitation.

Heartburn and dysphagia are not considered dyspeptic symptoms by the Maastricht consensus statement and this is now agreed internationally.

There are two broad categories of dyspepsia: organic and non-organic. Organic dyspepsia is produced by organic disease, e.g. peptic ulceration, oesophagitis, gastric carcinoma, etc., whereas non-organic dyspepsia is restricted to symptoms occurring in patients without any demonstrable focal lesion or pathology in the upper gastrointestinal tract. The prevalence of organic dyspepsia (symptoms and abnormal endoscopic changes) increases with age with a distinct threshold around 40–45 years. In particular, gastric cancer is rare below the age of 40 years.

Four subgroups of dyspeptic patients are recognized. These are based on the predominant symptoms and generally reflect the most likely aetiology of the symptoms:

- ulcer-like
- reflux-like
- dysmotility-like
- non-specific.

Although this grouping is clinically useful, the results of several prospective studies have demonstrated unequivocally that the symptoms alone are not able to differentiate between organic and non-organic causes of disease. In essence, conventional history taking is not reliably predictive of the underlying cause of the dyspepsia in the individual patient. The important practical issue in the management of patients with dyspepsia relates to which patients should be investigated by upper gastrointestinal endoscopy before treatment. In practice, there are a number of clinical variables that influence this decision, although practice tends to vary between centres. Thus patients with the following features generally have non-organic disease and are thus likely to be endoscopy negative:

- age < 45 years
- patients who are *H. pylori* negative
- patients without a history of NSAID usage
- patients without alarm/sinister symptoms: loss of appetite, weight loss and bleeding/anaemia.

Specific symptoms of surgical importance

Altered motility (primary or following gastric surgery) can result in rapid emptying of the stomach (dumping), delayed emptying (from gastroparesis) or abnormal reflux of duodenal contents into the stomach (enterogastric reflux). The symptom complex produced is varied but is determined by the functional abnormality. In some cases, the functional abnormality may in turn cause organic disease, e.g. gastritis from enterogastric efflux.

Loss of appetite, weight loss, recent-onset dyspepsia, constant upper abdominal pain and evidence of bleeding (overt or occult) have to be regarded as *alarm* or *sinister* symptoms, and thus require urgent investigation by endoscopy particularly if the patient is over 40 years of age. Weight loss is common, if not universal, in patients with postgastric surgery symptoms and is largely the result of reduced dietary intake, as in many of these patients the symptoms are precipitated by food or the patient is only able to eat small meals because of early satiety due to either reduced gastric reservoir or loss of the adaptive gastric relaxation after meals. Loss of appetite associated with early satiety/abdominal discomfort in a patient without previous gastric surgery is highly suspicious of an infiltrating gastric neoplasm and predates obstructive symptoms (vomiting). In most Western countries, pyloric obstruction with vomiting of ingested food not mixed with bile is very rarely caused by benign disease. In young female patients loss of appetite and weight is commonly caused by psychological disorders including anorexia nervosa.

Sudden, acute, severe constant epigastric pain accompanied by abdominal signs of peritoneal irritation implies a breach in the integrity of the gastric/duodenal wall, most commonly by peptic ulceration. Constant chronic back pain is indicative of a posterior ulcerating lesion (benign or malignant) penetrating the pancreas. In the case of duodenal ulcer, when this involves the head/neck of the pancreas, erosion of the gastroduodenal artery as it runs in the groove between the two organs can result in severe upper gastrointestinal haemorrhage.

Investigation of patients with gastric disorders

Endoscopy

The development of modern flexible fibreoptic endoscopes has allowed accurate diagnosis of both acute and chronic gastroduodenal disease, and in many instances altered its management. Undoubtedly, the experienced endoscopist has an accuracy rate superior to that of the radiologist. Its value in the bleeding patient is well established and, nowadays, it is axiomatic that all patients admitted with acute upper gastrointestinal bleeding require an upper gastrointestinal endoscopy within 24 hours of admission, and, when indicated, therapeutic control of the bleeding is achieved by this approach in the first instance.

The management of gastroduodenal disorders has changed radically with the recognition of the important pathogenetic role of infection with *H. pylori* and related organisms. This is reflected in the recommendation that testing for *H. pylori* should constitute the initial step in the management of patients below 45 years with dyspepsia and, if positive, eradication therapy commenced – *the test and treat strategy*, according to which endoscopy is avoided in patients who are rendered symptom free by this treatment. This approach has translated into considerable savings in endoscopy-related costs. This strategy has had other advantages. It has lowered the *H. pylori* prevalence and hence the risk of *H. pylori*-related disease including gastric cancer and reduced the endoscopic workload in most Western countries including the UK. However, this strategy is not without its problems. In the first instance, none of the tests for *H. pylori* are totally sensitive (10–20% false-negative rates). The *test and treat strategy* results in overtreatment of patients who may not require eradication, e.g. patients with reflux oesophagitis whose symptoms can be worsened by eradication and may miss significant diseases, e.g. Barrett's columnar cell change, gastric ulcer or worse still gastric neoplasm. In addition, it incurs a low morbidity related to the side effects of antibiotics and the emergence of resistant strains.

The *test and scope strategy* establishes the diagnosis before treatment, permits biopsy of organic disease and excludes gastric neoplasia. Furthermore, it reduces the anxiety of patients and avoids overtreatment. It is cost effective in Europe where the endoscopy costs (under 500 euros) are significantly lower than in the USA. The morbidity of upper gastrointestinal endoscopy is low but it is considered an unpleasant experience by most patients. In the current state of knowledge, the *test and scope strategy* is preferable, certainly in European countries. To a large extent, the validity of the alternative *test and treat strategy* relates to the prevalence of *H. pylori* in the community. In most Western countries, this has declined substantially during the past two decades and now is of the order of 10%, whereas high prevalence rates (40–70%) are still encountered in underprivileged countries. If the *test and treat policy* is adopted on the basis of high community prevalence, endoscopy should be done in dyspeptic patients who are either *H. pylori* negative or have alarm symptoms irrespective of their *H. pylori* status.

In light of the rather confusing situation, the European Panel on the Appropriateness of Gastrointestinal Endoscopy (EPAGE) considers upper gastrointestinal endoscopy to be *necessary* in the following:

- in individuals >45 years testing positive for *H. pylori*, with persistent symptoms despite eradication treatment
- in individuals >45 years, never previously investigated, *H. pylori* negative and no NSAID intake, with persistent symptoms despite acid suppression treatment
- in individuals >45 years with a previous history of gastric ulcer, no *H. pylori* testing or *H. pylori* negative, with persistent symptoms despite acid-suppression drug therapy.

Upper gastrointestinal endoscopy is a safe technique. Full facilities for resuscitation should always be available, and pulse oximetry is nowadays considered essential during the procedure. Sedation should always be kept to the minimum, particularly in patients who may have hypovolaemia or hypoxaemia. Diazepam or midazolam in small incremental doses is usually adequate. Fortral should not be used, as with any opiate it may depress respiration significantly, and it also delays gastric emptying, thus increasing the risk for subsequent regurgitation and aspiration. In addition, Fortral can cause pulmonary hypertension, which can be significant in patients recovering from haemorrhagic shock. Spasmolytic agents, such as Buscopan, are sometimes useful to allow visualization of the antrum, passage of the endoscope through into the duodenum and duodenoscopy.

Radiology

Barium meal examination is commonly used in investigating the stomach and duodenum. Although gross lesions are usually seen, the technique does have major limitations. Subtle mucosal changes of gastritis and of early mucosal cancer can only be detected if a meticulous double-contrast technique is used in which the stomach is filled with air and the mucosal lining coated with a thin film of barium. Even with these precautions, endoscopy gives superior results. A barium meal examination is notoriously unreliable in the assessment of patients with acute upper gastrointestinal bleeding. The presence of clot within the stomach produces a variety of bizarre appearances. Even when a lesion can be identified, contrast radiology cannot demonstrate whether this has been or is responsible for the bleeding. For these reasons, endoscopy is the preferred investigation in the bleeding patient. Despite these limitations, barium meal remains a valuable investigation in specific situations. A normal radiological appearance should not be accepted however in any patient with continuing dyspepsia and an additional endoscopy should be performed without delay. Barium swallow/meal is invaluable in outlining the exact topography of upper gastrointestinal cancer.

Endoscopic ultrasonography

Ultrasound examination of the stomach can be performed with either external or endoluminal probes. It is used to measure gastric wall thickness, to assess the extent of intramural involvement of the stomach wall by tumour and to establish enlargement of the perigastric lymph nodes. The normal stomach wall thickness measured by ultrasound (internal to external) is 5–6 mm. An endoscopy with biopsy is indicated when the ultrasound gastric wall thickness exceeds 5 mm.

High-frequency flexible endoscopic ultrasonography (EUS) is now well established in routine practice and distinguishes five layers of the stomach wall: superficial mucosa, deep mucosa, submucosa, muscularis externa and serosa/adventitia (Figure 23.3). It provides the most sensitive modality for the accurate staging of gastric tumours, being superior in this respect to all other imaging tests. In the stomach, EUS is particularly useful in:

- the pretreatment staging of early gastric cancer in detecting absence, superficial and deep involvement of the submucosa

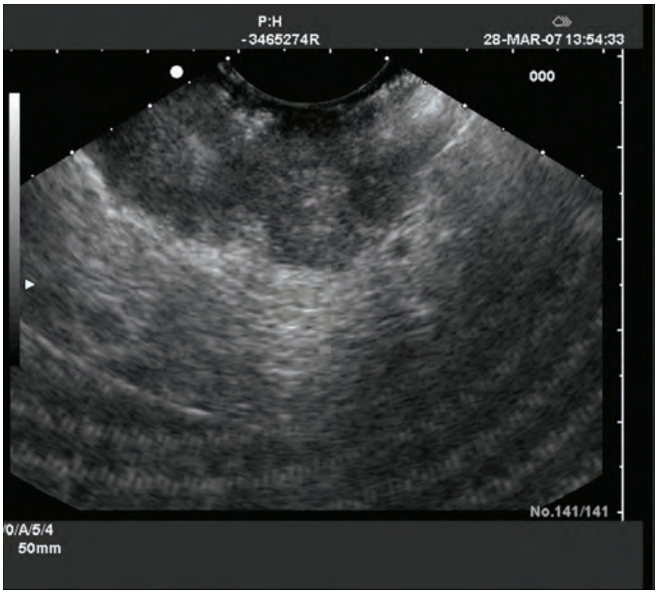


Figure 23.3 Endosonography of a gastric cancer.

- the staging and follow-up of MALT lymphomas
- the diagnosis of stromal (mesenchymal) tumours.

EUS is the established diagnostic modality of choice in local (T) staging of gastric cancer. The alternatives are multidetector row CT (MDCT) and MRI. A recent systematic review compared the relative performance of the three imaging modalities. The overall diagnostic accuracy of T staging for EUS, MDCT and MRI was similar, varying among 65–92.1%, 77.1–88.9% and 71.4–82.6%, respectively. The sensitivity for detection of serosal involvement for EUS, MDCT and MRI was also not statistically different: 77.8–100%, 82.8–100% and 89.5–93.1%, for EUS, MDCT and MRI, respectively, with corresponding specificities for the detection of serosal involvement of 67.9–100%, 80–96.8% and 91.4–100%. Thus although all three modalities give similar results in T staging and in assessing serosal involvement, EUS remains the first-choice imaging modality in preoperative T staging of gastric cancer.

Deep involvement of the submucosa in early gastric cancer (EGC) is associated with a significant increased risk of lymph node involvement. Thus whereas EGCs limited to the mucosa have a 3% incidence of regional node involvement, this increases to 20% with deep involvement of the submucosa, thereby precluding any form of local endoscopic treatment. EUS is also capable of identifying advanced gastric cancer (involvement of the muscularis propria) and infiltration of adjacent organs, but is unreliable in detecting lymph node involvement (as opposed to enlargement) and peritoneal deposits.

EUS has an overall diagnostic accuracy for gastric non-Hodgkin's lymphoma, approximating 95%. It is also reliable in assessing depth of invasion (85% accuracy) and, to a lesser extent (60%), in detecting metastatic spread to the perigastric lymph nodes. The characteristic echo pattern consists of an extended hypoechoic thickening of the second and third EUS layers with preservation of the identity of all the layers. A similar picture is obtained in linitis plastica except that the hypoechoic thickening is more longitudinally extensive and involves the entire circumference of the stomach. In patients with MALT lymphomas associated with *H. pylori* infection, if the lesion is confined to the second or third layer on EUS staging (deep mucosa or submucosa), eradication of the infection results in remission in 85% of cases. In contrast if the EUS staging shows deeper infiltration, eradication of the *H. pylori* infection has no effect on the tumour and indicates the need for chemotherapy or surgery.

Mesenchymal (stromal) tumours appear as hypoechoic masses continuous with the fourth (muscularis externa) and fifth (serosa/adventitia) layers. The EUS features suggestive of malignancy are tumour size >4 cm, irregular extraluminal border, echogenic foci and cystic spaces. If all these features are absent, the stromal tumour can be predicted as benign. However, a confident diagnosis of malignancy should not be made on EUS findings alone.

Multidetector row CT

There is a confusing array of different names in use for this imaging technology: multidetector row CT (MDCT), multidetector CT (MDCT), multidetector array helical CT, multichannel CT and multislice CT (MSCT). The number of simultaneous measurements along the length of the patient's long axis is referred to as the number of 'slices', e.g. 64-slice MDCT. Aside from the extremely high resolution, MDCT enables volumetric three-dimensional (3D) visualization and is used routinely in the preoperative staging of gastric cancer with or without contrast enhancement.

Studies comparing two-dimensional (2D) vs 3D MDCT imaging for gastric cancer indicate improved T staging of gastric with the latter (92% vs 77%). However, the overall accuracy rate for N staging is not improved by 3D volume visualization. The overall detection rate for gastric cancer by MDCT averages 90%, with detection rates for advanced gastric cancer approaching 100% and 85% for early gastric cancer. MDCT-based virtual gastroscopy is used in some centres for the characterization of both advanced and early gastric cancers. In a volumetric MDCT study from Japan on the depth of invasion involving scanning of 41 patients after drinking 600 mL of water and injection of non-ionic contrast agent, the overall sensitivity, specificity and accuracy rate of MDCT in determining the invasion of serosa were 85%, 96% and 92%, respectively.

2D MDCT is the standard test in most hospitals that is used in the preoperative staging and assessment of operability of patients with gastric cancer. Some centres use routinely hydro-MDCT. This technique involves the ingestion of water or dilute contrast (600–1000 mL) followed by scanning with the patient in the prone position (Figure 23.4). Hydro-MDCT allows documentation of the extent of mural involvement, extragastric extension and involvement of the retrogastric organs such as the pancreas. MDCT also enables detection of lymph node enlargement especially of the coeliac/para-aortic lymph nodes.

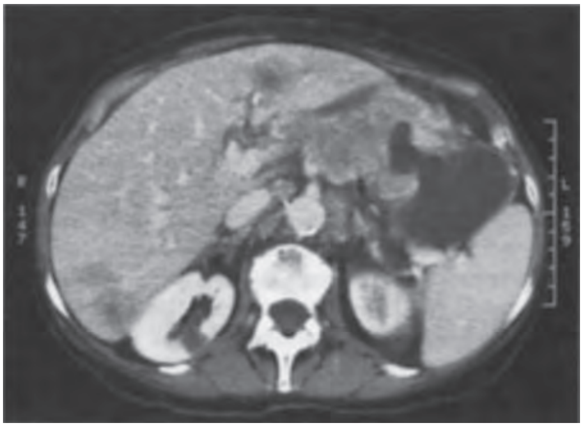


Figure 23.4 Hydro-multidetector row CT in a patient with advanced gastric cancer showing an advanced T3 tumour with local extragastric spread.

By contrast CT is not as reliable as other modalities in the staging of primary gastric lymphomas in terms of both extent of mural invasion and regional node involvement. It is also poor at distinguishing low- from high-grade disease. Hence, other modalities especially high field (3.0 T) MRI and EUS are preferred.

Magnetic resonance imaging

Despite, the considerable advances in MRI (3–7 T systems) with advanced software capable of 3D reconstruction, dynamic imaging, etc., MRI is not currently the gold standard (currently CT and EUS) in the preoperative staging of gastric cancer and there have been few published comparative studies between MDCT and MRI in the staging of advanced gastric cancer (invading the muscularis propria). One such study, prospective in nature, involved examinations in 26 patients with biopsy-confirmed disease. Contrast-enhanced MDCT and non-enhanced MRI with a 1.0 T machine using various sequences were obtained in each patient following injection of antiperistaltic drug and ingestion of 1.0 L of tap water. The images obtained by the two modalities were analysed by two radiologists. Diagnostic accuracy of each staging by MDCT or MRI was evaluated by comparison with the pathological results obtained from the resected specimen (histological staging). In this study MRI was slightly but significantly superior to CT in T staging, although it had a tendency to overstage the pathologica T2 cancers. With MRI, both the positive predictive value for T2 stage and sensitivity for T3 stage reached 100%. Regarding the N staging, CT was slightly superior to MRI (73% vs 65%; $p > 0.05$). However, both the two modalities tended to understage the N staging. At the time of writing, there are no published series of MRI staging of gastric cancer with 3.0 T machines, which provide a significantly higher resolution. The prediction is that high-field MRI systems will take over from MDCT in time, especially as they do not expose the patients to ionizing radiation.

Positron emission tomography CT

This medical imaging system combines positron emission tomography (PET) with MDCT to enable functional (metabolic activity) imaging with 2D/3D anatomical imaging of the disease as the two images are superposed (co-registered) (Figure 23.5). The combined systems share common software for the 2D/3D reconstruction. PET is a functional nuclear medicine imaging system which detects pairs of gamma rays emitted by a positron-emitting radionuclide administered to the patient before the imaging. The radionuclide most commonly used is a metabolite of glucose, ^{18}F-fludeoxyglucose (^{18}F-FDG or FDG). This contains the positron-emitting radioisotope fluorine-18 at the 2 hydroxyl ($^{-}$OH) group of the glucose molecule. The 2 hydroxyl group ($^{-}$OH) in glucose enables the metabolism by glycolysis of this nutrient. Hence although ^{18}F-FDG is taken up by active metabolizing cells, it is not capable of being metabolized and furthermore cannot exit the cell, disappearing only from the cellular tumour mass with radioactive decay. Thus, it provides imaging of living actively metabolizing tumour cells. Although the radioactivity elimination half-life of ^{18}F-FDG is the same as that of fluorine-18 (110 minutes) clinical reports on its use have shown that only 80–85% of the fluorine-18 activity is eliminated with a half-life of 110 minutes, as the rest is eliminated rapidly by the kidneys soon after injection (half-life of 16 minutes). This is important as the urine of a patient undergoing PET remains radioactive for several hours after the examination and adequate precautions are needed to eliminate the risk of contamination of bed linen, etc. One of the practical limitations of both PET and PET-CT relates to the production and transport of the radiopharmaceutical needed for PET

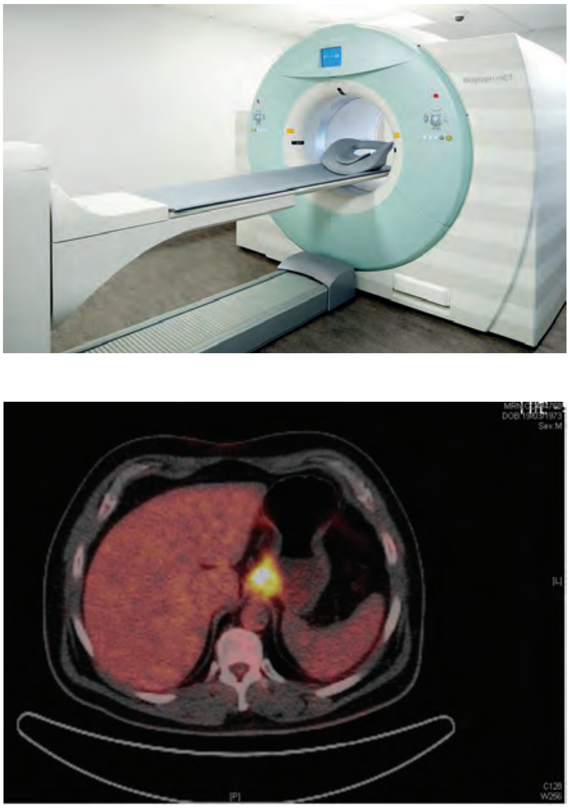


Figure 23.5 (a) Positron emission tomography (PET)-CT system; (b) positive PET scan.

imaging as these are extremely short lived and the hospital has to be within easy access to a cyclotron which is used for their production. PET-MRI systems have been introduced in recent years and may well replace PET-CT systems, especially with the advent of high-field MRI scanners.

To date, there have been few published series on the use of PET-CT in gastric cancer. However, early reports suggest its usefulness and superiority over MDCT alone in the diagnosis of recurrence after surgery for gastric cancers. In one series from Korea on 88 patients, PET/CT detected 40 locoregional lesions, 29 retroperitoneal and pelvic lesions, 25 seeding and 12 distant metastases. CT with contrast enhancement detected 28 locoregional lesions, 20 retroperitoneal and pelvic lesions, 27 seeding and 12 distant metastases. However, false-positive results were observed with PET–CT scanning in 23 patients (20%).

Other investigations

In most cases the morphological assessment of the stomach and duodenum by contrast, the various imaging modalities and endoscopy give sufficient information for further management decisions to be made. In selected cases, it may be necessary to obtain a functional evaluation of the gastroduodenal region. This can be achieved by gastric secretory testing, measurement of gastric emptying and enterogastric reflux, and determination of the plasma concentration of the antral hormone, gastrin.

Gastric secretory tests

Because of the association of duodenal ulceration with hypersecretion of acid, measurement of gastric acid output has held a fascination for many surgeons. In practice, however, measurement of acid secretion has little clinical relevance in the management of patients with upper gastrointestinal disorders unless Zollinger–Ellison syndrome is suspected. Thus nowadays and especially with the drastic decline in the use of vagotomy for the treatment of duodenal ulceration, acid secretory studies are carried out largely for research purposes.

Basal and maximal acid output

Basal secretion is usually measured over 1 hour. Maximal secretion is best stimulated with the synthetic gastrin analogue, pentagastrin, which is administered by intramuscular injection in a dose of 6 g/kg (10 pg/kg in the postvagotomy patient). Secretion is then followed for 1 hour and the maximal response (peak 15 minutes ×4) calculated. There are many pitfalls in acid secretory tests. The overlap between normal and disease is considerable, and the results obtained vary with body weight and sex. Typical values in preoperative and postoperative patients are shown in Table 23.1. The only clinical indication for acid secretory studies nowadays is in the diagnosis of pernicious anaemia and Zollinger–Ellison syndrome.

In the postoperative patient, maximal acid output is reduced by 70% or more after gastrectomy, and by 50–70% by vagotomy. A postoperative maximal acid response of 20 mmol/h is a fairly reliable indicator of a persisting hypersecretory state (incomplete vagotomy).

The measurement of basal acid output after gastric surgery has particular relevance to the question of ulcer recurrence caused by a pancreatic gastrin-producing tumour (Zollinger–Ellison syndrome). If hypergastrinaemia exists, the 'basal secretion' is already being stimulated and the characteristic findings will be of a high basal secretion, which does not increase much further on maximal stimulation. If basal secretion is greater than 10 mmol/h, and this represents more than 60% of the maximal response, a gastrinoma should be suspected. Proof of this requires measurement of plasma gastrin concentration.

The insulin test

After complete denervation of the parietal cell mass, acid output is reduced by 50% or more. Incomplete denervation will not achieve this, and the risk for a recurrent ulcer is increased. Nowadays, an assessment of the adequacy of vagotomy is indicated only in *H. pylori*-negative patients who develop recurrent ulceration after this procedure. The insulin test, originally described by Hollander, depends upon the fact that insulin-induced hypoglycaemia stimulates hypothalamic nuclei which induce a parasympathetic response. If vagal innervation of the parietal cell mass persists, acid will be secreted in response to the hypoglycaemia. This test is performed by giving an intravenous injection of soluble insulin (0.2 U/kg) after the basal secretion has been collected in four 15 minute aliquots. Secretion is then collected for a further 2 hours. If the acid concentration in any 15 minute sample after insulin is 20 mmol/L greater than in the basal period, the test is regarded as positive. If no free acid is secreted basally, then a postinsulin concentration of more than 10 mmol/L is taken as positive. Subsequent to Hollander's original description, the interpretation of the test has been modified in many ways. The modification most commonly used is to categorize a positive response as *early* (during the first hour) or *late* (second hour).

A positive insulin test indicates residual vagal innervations of the parietal cell mass. An *early* positive response usually implies fairly generous residual innervations and the actual amount of acid secreted (i.e. volume as well as concentration) is quite large. If a patient with recurrent ulcer has an *early* positive response to

Table 23.1 Typical values of basal and maximally stimulated acid secretion

	Acid output (mmol/h*)	
	Basal secretion	Maximal secretion
Normal	2	20–30
Preoperative duodenal ulcer	>5	>35
Postvagotomy	<2	10–20
Postgastrectomy	1	<10

*These values are approximations only. Secretion increases with increasing body weight and decreases progressively with age. Females secrete about two-thirds the amount of acid secreted by males.

insulin and a maximal acid output of 20 mmol/h or more, then incomplete vagotomy is the cause. A *late* positive response may not be so significant, particularly if the acid output in response to insulin and pentagastrin is small. Cephalic stimulation can also be produced by sham feeding (chew and spit). The gastric secretory response to sham feeding is used in some centres instead of the insulin test, as it is safer and just as reliable. One of the problems with the insulin test is that many asymptomatic patients can show a positive response. Furthermore, when serial tests have been performed after vagotomy, it has been shown that negative responses in the first few months after operation can revert to positive later on. Thus the insulin test must be interpreted with caution, and considered only in the context of all other clinical, endoscopic and secretory information. The deliberate production of hypoglycaemia is not without problems, and the test should not be carried out in patients with significant cardiac disease, cerebrovascular disease or epilepsy.

Gastric emptying

Measurement of gastric emptying is indicated in patients with symptoms indicative of dysmotility-type dyspepsia. The two conditions which may arise spontaneously (consequent on systemic disease) or develop after surgery on the upper gastrointestinal tract are:

- gastroparesis – early satiety, fullness and epigastric retention with vomiting
- rapid gastric empting with early vasomotor dumping.

In surgical practice, the majority of these patients have postgastric surgery symptoms and are investigated with a view to possible remedial surgery. It seems likely that future sufferers are more likely to be patients who have undergone bariatric gastric bypass surgery. Delayed gastric emptying can also be encountered after fundoplication (open or laparoscopic) for gastro-oesophageal reflux disease, especially patients who have been subjected to redo procedures for recurrence of reflux or for some other complication related to the fundoplication. Delayed gastric empting can also be the result of primary dysmotility disorders including the rare *hollow visceral myopathy*, which affects the entire gastrointestinal tract, and the much more common visceral neuropathy, which may complicate longstanding diabetes mellitus.

A variety of radionuclide meals are used for studying gastric emptying in patients with symptoms suggestive of gastroparesis or vasomotor dumping. The patient drinks or ingests the radionuclide-labelled meal and the radioactivity in the stomach is then monitored with an external gamma camera, thus obtaining radioactivity–time curves. The most useful index is the $T_{1/2}$, which is the half-emptying time, i.e. the time taken for half of the ingested meal to leave the stomach. Assessment of both liquid and solid gastric emptying (on separate occasions) is necessary to document the pattern of abnormal gastric emptying in the individual patient. Thus in many postgastrectomy patients, liquid emptying is fast but solid emptying is delayed (Figure 23.6). Rapid gastric emptying of liquids is often associated with a much shorter small bowel transit time. Useful qualitative information can be obtained using a 'physiological' food and barium meal if the more sophisticated isotope techniques are not available.

Test for enterogastric reflux

Measurement of gastroduodenal reflux can assist in the management of patients with bile vomiting and bile gastritis. In the latter case, demonstration of increased quantities of duodenal contents, including bile refluxing into the stomach, may suggest that reoperation for bile diversion may improve the gastritis and the patient's symptoms. Enterogastric bile reflux is best quantified by combining biliary excretion scintigraphy (HIDA scan) with CCK or preferably a milk meal (as a source of fat) – the *milk-HIDA test* (Figure 23.7).

Plasma gastrin

The peptide hormone gastrin is elaborated in specific endocrine G-cells in the gastric antrum and, to a lesser extent, the duodenum. Gastrin secretion occurs in response to distension

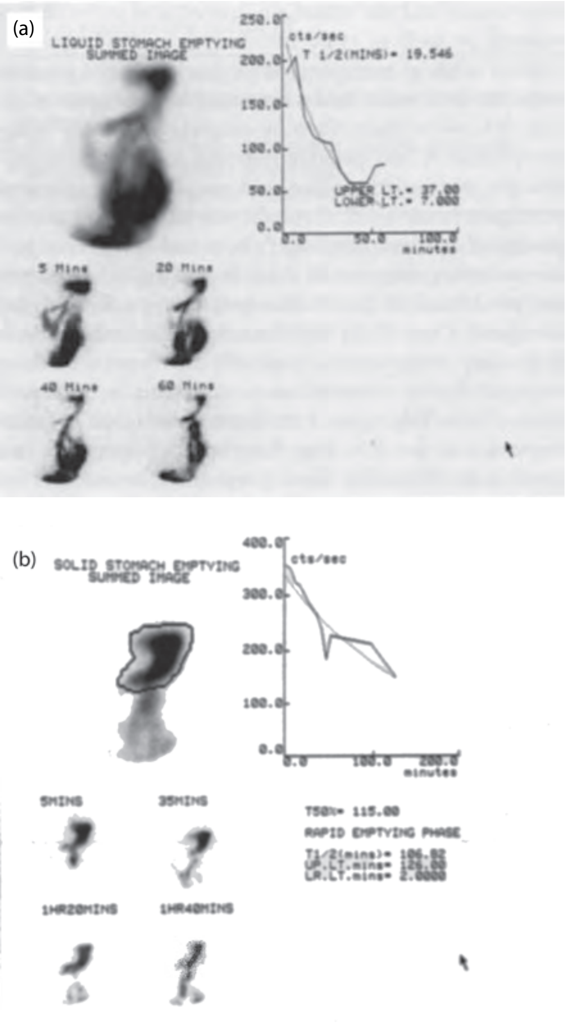


Figure 23.6 Isotope gastric emptying in a patient with vasomotor dumping after truncal vagotomy and drainage: (a) liquid meal showing fast emptying; (b) solid meal demonstrating slow emptying.

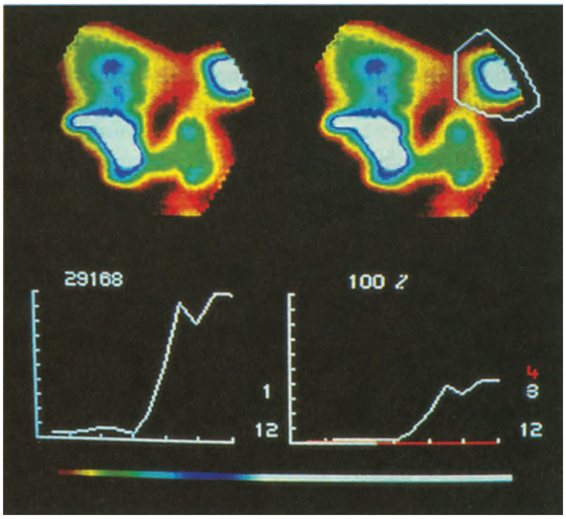


Figure 23.7 Extensive reflux of bile into the gastric remnant in a patient with severe bile gastritis and vomiting after Pólya gastrectomy.

of the antrum in the presence of food, particularly protein. When the pH in the antrum falls, gastrin release is inhibited by somatostatin (produced by the D-cells), such that this negative control feedback prevents prolonged gastric hyperacidity in the normal state. The concentration of gastrin in the plasma can be measured by specific radioimmunoassay. This is indicated if clinical and/or gastric secretory features suggest the possibility of Zollinger–Ellison syndrome. The reported levels of plasma gastrin vary between laboratories, but values greater than 200 pg/mL are regarded as abnormally high. Values in excess of 1000 pg/mL are virtually diagnostic of a gastrinoma, provided that the patient is secreting acid. The reason for this proviso is that in any achlorhydric patient, e.g. one with pernicious anaemia, there is no fall in the antral pH to shut off gastrin secretion by somatostatin and hence a persistent hypergastrinaemia ensues, although this does not normally reach the high levels associated with a gastrinoma.

In some patients, the plasma gastrin concentration will be found to be elevated but not sufficiently for the diagnosis of Zollinger–Ellison syndrome to be confirmed. There are various causes of hypergastrinaemia (Box 23.1). Many of these can be excluded readily on clinical grounds, but difficulty may be experienced in distinguishing a gastrinoma from antral G-cell hyperplasia, retained excluded antrum or merely an exaggeration of the usual rise in gastrin that follows acid reduction by vagotomy. Three stimulatory tests are useful in this respect: protein meal, calcium stimulation test and secretin challenge test (Figure 23.8).

BOX 23.1 Causes of hypergastrinaemia

Primary autonomous hypersecretion
- Tumour (gastrinoma) – Zollinger–Ellison syndrome
- Antral G-cell hyperplasia

Increased stimulation
- Hypercalcaemia

Decreased inhibition
- Hypo- or achlorhydria – pernicious anaemia, post vagotomy
- Retained and excluded gastric antrum
- Small bowel resection

Decreased removal
- Renal failure
- Small bowel resection

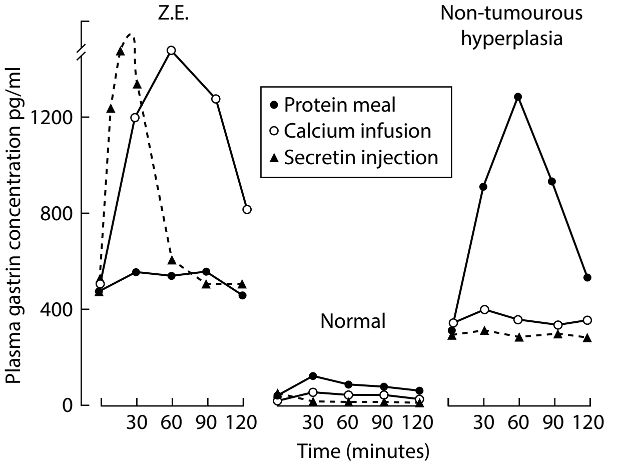


Figure 23.8 The use of provocative gastrin stimulation tests to differentiate between the hypergastrinaemia of the Zollinger–Ellison (ZE) syndrome and that due to antral G-cell hyperplasia.

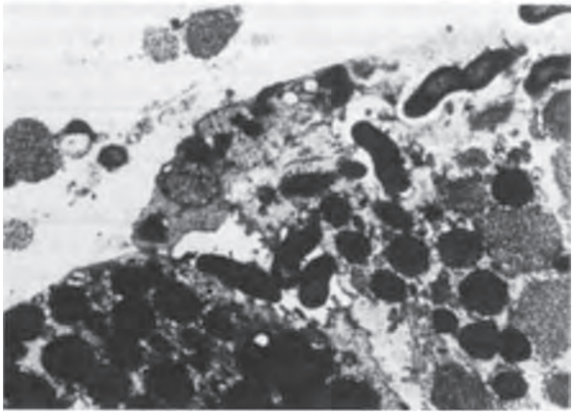


Figure 23.9 *Helicobacter pylori* in a gastric mucosal biopsy.

Helicobacter pylori infection

Gastroduodenal disease

H. pylori, previously known as *Campylobacter pylori*, is a spiral-shaped organism (Figure 23.9) of great aetiological importance in certain gastroduodenal disorders. The infection is acquired by oral ingestion. Transmission by inadequately cleaned and disinfected endoscopes is now well documented. There is also evidence of transmission between cohabiting parties as spouses/partners of duodenal ulcer patients have a higher seroprevalence of *H. pylori* infection than controls. The organism is often identified in children complaining of acute abdominal pain. Indeed, population studies indicate that the infection is usually acquired in childhood, and thereafter the acquisition rate slows down exponentially with age. A high rate of infection is documented among endoscopy staff (endoscopist and nurse working in endoscopy units), with prevalence rates as high as 75–80% compared with 20–65% background infectivity rates (varies considerably in different countries). Indeed in most Western countries *H. pylori* infection is fast disappearing, but infection rates remain high in developing countries.

Pathology

Within the stomach, *H. pylori* localizes on the epithelial surface beneath the viscid mucous layer, where it exerts its pathological effects via the elaboration of various enzymes and toxins. The initial acute inflammation persists as a chronic inflammation which predominantly affects either the antrum or the body of the stomach with different consequences in terms of disease: antral disease causes duodenal ulcers in patients with a large parietal mass, whereas body-predominant disease causes gastritis, gastric atrophy and gastric cancer.

The gastric mucous barrier is disturbed by the production by *H. pylori* of *endopeptidase*, which has a powerful mucolytic action, and by the generation of large amounts of ammonia with an increase in the epithelial pH. This alters the mucosal charge gradient, cellular permeability and epithelial Na^+, K^+–ATPase activity, leading to back diffusion of H^+. The adhesion of the organism to the surface of the gastric epithelial cells is mediated by a specific substance, *adhesin*, which also causes haemagglutination of erythrocytes *in vitro*. In addition to surface attachment, the organism can invade the gastric epithelial cells. *H. pylori* produces two cytotoxins. One of these induces vacuolization of the affected cells. This *vacuolizing toxin* is much more commonly encountered in isolates from patients with duodenal ulceration than those with gastritis and no ulceration. The clinical relevance of the other toxin, which is a weak haemolysin, remains unknown. The other adverse effects of *H. pylori* infection include reduced epithelial cell turn over (impaired healing), reduction of surface hydrophobicity and increased lipolytic activity, release of chemotactic factors for neutrophils and monocytes and degranulation of eosinophils with release of cytotoxic cationic proteins, increased apoptosis, activation of the classical complement pathway and stimulation of the G-cells with increased release of gastrin as a result of the high surface pH consequent on the locally produced NH_3.

Not all patients with *H. pylori* infections develop gastroduodenal disease. Aside from host factors, the virulence and the genetic type of the organism, age of the host and environmental factors (smoking, diet especially the protective effect of dietary fruit and vegetables) are important and determine the development of specific gastroduodenal disease, depending on the predominant site of the infection – antrum or body – and its incidence. *H. pylori* exhibits marked genetic variation and hence the presence of virulent and less virulent strains.

Virulence is associated with the *cag* pathogenicity island (PAI) (a segment of DNA) which has 40 genes, one of which is the *cytotoxin-associated gene A (cagA)* that codes for the protein CgA that is highly antigenic and hence can be detected by antibody tests. The PAI is associated with the vacuolating toxin and hence mucosal damage and inflammation. Longstanding inflammation then leads to atrophy and intestinal metaplasia. The *H. pylori* antigens also induce the appearances of MALT follicles in the gastric submucosa. Hence the entire spectrum of gastroduodenal disease caused by *H. pylori* is gastroduodenitis, duodenal and gastric ulcers, atrophic gastritis and gastric cancer, and MALT gastric lymphomas.

What is interesting is that the PAI including the *cagA* gene does not produce the vacuolating toxin. This is coded by another gene (*vacA*) that is located in the genome of *H. pylori* outside the PAI. However, only organisms that possess the PAI can secrete the vacuolating toxin. Whereas infections with virulent *CagA*-positive organisms cause serious gastroduodenal disease, paradoxically they protect the host from the development of reflux oesophagitis and probably Barrett's columnar metaplasia and carcinoma of the gastro-oesophageal junction. Indeed the marked fall in the incidence of *H. pylori* infections in Western countries has been linked with the marked rise in the incidence of these junctional cancers over the past 20 years.

Extradigestive disease

There are several mechanisms whereby gastric infection with *H. pylori* can contribute to the development of extradigestive disease:

- immunological responses to *H. pylori* antigens
- constant production of proinflammatory substances: cytokines, eicosanoids, acute-phase proteins
- mimicry between bacterial and host antigens
- reduced folate and iron absorption.

This has generated past and continued research interest into the possible aetiological association between *H. pylori* infection and a variety of extradigestive disorders (Box 23.2).

The reported studies in patients with ischaemic heart disease have tended to confirm this association, but with an extremely wide variation on the estimates of the degree of risk of infected patients. One of these reviews has indicated that infected patients carry a low (10–20%) excess risk. The association between

BOX 23.2 Extradigestive disorders with possible association with *H. pylori* infection

- Ischaemic heart disease
- Ischaemic cerebrovascular disorders, including carotid artery stenosis
- Functional vascular disorders: Raynaud's phenomenon and idiopathic migraine
- Immunological disease: Henoch–Schönlein purpura, Sjögren syndrome, autoimmune thrombocytopenia
- Extragastric MALT lymphomas
- Sideropenic anaemia: possible iron malabsorption
- Skin disorders: idiopathic chronic urticaria, acne rosacea and alopecia areata
- Hepatobiliary disease: cirrhosis, cholesterol stones. Chronic cholecystitis (other *Helicobacter* species, e.g. *H. bilis*, *H. pullorum*)

H. pylori infection and carotid artery stenosis is based on reports that show significantly more severe stenosis in infected than in non-infected patients suffering from this condition. The hypothesis suggested for the role of *H. pylori* infection in atherosclerosis, aside from the chronic inflammatory cascade, is folate malabsorption by the infected stomach. Low serum folate is a well-documented risk factor for ischaemic vascular disease.

Helicobacter heilmannii infections

This organism, previously known as *Gastrospirillum hominis*, can also infect the stomach, although it is much less commonly encountered in clinical practice and is reported to be present in 0.08–1% of endoscopies. Infection produces a gastritis that differs from that caused by *H. pylori* in several respects. In the first instance, the mononuclear infiltrate is mild, the inflammation is usually inactive, the bacterial mucosal load is low and the degenerative/regenerative lesions in the surface gastric epithelium are focal and mild. Contrary to *H. pylori*, the bacteria do not adhere to the surface epithelium. Nonetheless, infection with *H. heilmannii* can cause peptic ulcer disease, gastric cancer and gastric MALT lymphomas. Thus when present the condition requires eradication therapy.

Tests for *Helicobacter pylori* infection

The most commonly performed are the *rapid urease tests*. These are carried out on endoscopic biopsies and use test kits such as CLO, Hfast, Pyloritec, etc., and provide a result within 3 hours of the endoscopy. The sensitivity and specificity of these tests at the final reading are similar at 90% and 100%, respectively. Therapy with PPIs at the time of endoscopy significantly reduces the sensitivity of antral and corpus biopsies for detection of infection by rapid urease tests and histology. Thus, if PPI therapy cannot be discontinued, the *rapid urease test* should not be read before 24 hours after biopsy, and, in addition, a serological test should be performed.

Other tests include culture in a microaerobic environment, the polymerase chain reaction, histology of antrum and corpus (Giemsa or Warthin–Starry silver stain), the ^{13}C breath test and serology for the detection of *H. pylori*-specific antibodies [immunoglobulin (Ig) C]. The modern immunological test kits are based on purified antigen preparations and use the enzyme-linked immunosorbent assay.

There is little to choose between the various tests in terms of sensitivity and specificity (Table 23.2) and in practice the choice of test depends on the clinical situation. In symptomatic patients undergoing endoscopy, a biopsy-based test is the appropriate one (*rapid urease*, histology, culture) and the accuracy is increased if more than one test is performed. In asymptomatic patients a serological test is first performed, and if this is positive confirmation is achieved by the ^{13}C breath test. This test is also indicated in patients with recurrent symptoms after previous eradication, even if the endoscopy is normal. Serological tests within 6–12 months of eradication are unreliable. It is recommended that the outcome of any eradication treatment needs to be confirmed by a non-invasive test 4–8 weeks after the end of treatment, and the best test for this purpose is the ^{13}C breath test.

Table 23.2 Comparative sensitivity and specificity of the various tests used for the diagnosis of *Helicobacter pylori* infections

Test	Sensitivity (%)	Specificity (%
Rapid urease tests	90	100
Culture	98	100
Polymerase chain reaction	97	100
Histology of antrum and corpus	98	99
^{13}C breath test	100	100
Serology	98	88

Acute gastroduodenal disorders

The various surgical disorders of the stomach and duodenum are discussed later in this chapter. Many of these diseases present with varying degrees of urgency, such as perforation, bleeding or stenosis. The general management of these situations is discussed here.

Perforation

Free perforation into the general peritoneal cavity is a catastrophic event, the signs and symptoms of which do not usually cause diagnostic problems. When perforation of a chronic peptic ulcer occurs, there will often have been an increase in the severity of the dyspepsia for a few days prior to the perforation. When an acute ulcer perforates, there may be no premonitory symptoms, particularly in younger patients. The pathology of steroid-induced perforations is not fully understood. For instance, perforation is much more common in patients with rheumatoid arthritis on steroids than in patients with ulcerative colitis. This may be a reflection of the primary disease process, or, alternatively, may be a feature of the concomitant use of other drugs. The use of NSAIDs has been shown to be associated with an increased risk of both perforation and acute bleeding.

Acute perforations also accompany situations of stress, such as burns, multiple injuries and sepsis. It can also occur in patients receiving intensive chemotherapy and radiotherapy for malignant disease. Such perforations may be duodenal, pyloric or gastric. Perforation of malignant gastric ulcers is common and although most perforated gastric ulcers will be benign, biopsy is necessary if the ulcer is not removed.

The moment of perforation is often identified by the patient as sudden excruciating epigastric pain. The subsequent symptoms depend in part on the degree of peritoneal soiling and whether the perforation becomes sealed by the greater omentum. Otherwise, the pain becomes generalized. In addition, the patient may experience referred shoulder pain caused by diaphragmatic irritation. Sometimes, the spread of the gastroduodenal contents is maximal along the right paracolic gutter, such that the intensity of the pain becomes localized in the right iliac fossa, thus simulating acute appendicitis. Significant vomiting is uncommon unless the diagnosis is delayed and a paralytic ileus becomes established. Occasionally in elderly or seriously ill people (usually in intensive care), the perforation does not develop as a dramatic episode and there is a more gradual development of generalized peritonitis with fewer intense symptoms and signs.

The physical signs accompanying perforation will again depend upon the degree and rate of peritoneal soiling. Tenderness, with guarding, may vary from being localized to the upper abdomen to being generalized. If contamination of the general peritoneal cavity has occurred, there will be marked (board-like) rigidity, rebound tenderness and a silent abdomen. Abdominal distension is a late feature and reflects the onset of paralytic ileus and occurs because of subsidence of the reflex spasm of the anterior abdominal musculature. In elderly, immunosuppressed and seriously ill patients, this may be the only significant clinical finding. A variable degree of peripheral circulatory failure may be present, with tachycardia, hypotension, a cold periphery and a decreased urinary output. Respiration will be shallow and grunting.

Diagnosis

The key to the diagnosis, suspected from the symptoms and signs, is the plain abdominal or chest radiograph taken in the erect position. The lateral decubitus position may be used instead in a very ill patient. Although free peritoneal gas may come from any of the alimentary hollow organs, in practice the finding of subdiaphragmatic air usually on the right side is virtually pathognomonic of gastroduodenal perforation. Radiological features of an ileus may be present in more advanced cases. If a pneumoperitoneum is not evident radiologically, the diagnostic problem concerns the differentiation between a sealed perforation with minimal localized soiling and acute pancreatitis. In some instances of the former, identification may not be important if the clinical state of the patient is stable or improving. Otherwise, differentiation between a sealed perforation and acute pancreatitis is obtained by contrast radiology and estimation of serum amylase. Although moderate elevation of the serum amylase may be present in 10–20% of perforated ulcers, it is uncommon for this to exceed 700 Somogyi units unless renal function is impaired as a result of the hypovolaemia. If doubt persists, a diagnostic peritoneal tap or lavage may help. The amylase content will usually be very high in either condition, but it should be possible to differentiate between gastroduodenal contents and the brownish fluid of acute pancreatitis. An alternative approach is diagnostic laparoscopy, but this requires a general anaesthetic. If this goes ahead and the findings are those of a sealed duodenal ulcer, no attempt should be made to disrupt the omental seal. Aspiration of any fluid and saline irrigation is all that is required.

Management

Initial treatment should be directed towards correction of hypovolaemia and any electrolyte deficit/imbalance. Oliguria and poor peripheral perfusion are contraindications to immediate operative treatment, and their correction should take precedence even over radiological studies. If necessary, resuscitation should be monitored with measurement of central venous pressure and urine output from an indwelling urinary catheter. Colloids can be used for resuscitation, but crystalloids in the form of balanced electrolyte solutions are equally effective. Oxygen via a mask is recommended in shocked patients. Pain relief should be given immediately even before the physical examination. Intramuscular pethidine is usually very effective, and a small dose can be given intravenously if absorption from a poorly perfused periphery is suspected in the presence of hypovolaemia. Nasogastric intubation with intermittent aspiration is instituted next. Antibiotics are not recommended routinely since the initial peritonitis is chemical. However, if the presentation is late or surgical treatment is delayed beyond 8 hours from the onset of symptoms, or if the patient has chronic respiratory problems, the use of a broad-spectrum antibiotic (usually a second-generation cephalosporin) is justified.

Operative or conservative management

There is no doubt that some patients with perforated peptic ulcers can be managed conservatively (non-operatively) with a successful outcome with the understanding that these patients will be investigated by upper gastrointestinal endoscopy on recovery from the acute illness. A policy of nasogastric suction, intravenous fluids, antibiotics and analgesics will allow many perforations to seal spontaneously and the ileus to resolve. An H_2-blocker or more recently a PPI (esomeprazole, pantoprazole, etc.) administered intravenously is used by some to lower gastric secretion, gastric pH and juice volume. The problem with the conservative approach, which is usually reserved for poor-risk elderly patients, is a high incidence of residual abscess formation, particularly in the subphrenic regions, which will subsequently require drainage, although this may be carried out percutaneously under radiological guidance, thus obviating surgical intervention.

An operative approach (closure of the perforation and peritoneal lavage) is therefore strongly recommended in the vast majority of patients. This may be performed through an open surgical or laparoscopic approach. Several retrospective studies and two randomized controlled trials have confirmed that the clinical outcomes of patients in terms of survival and morbidity between the two approaches are identical. However, the laparoscopic approach has several disadvantages, including less postoperative pain with reduction in the analgesic intake and faster recovery.

Open surgical treatment

This is usually performed through an upper midline incision and consists of closure of the perforation with interrupted sutures (small duodenal perforation) or an omental patch (large duodenal perforation or gross friable oedematous perforation). For perforated gastric ulcers (other than juxta-pyloric) excision of the ulcer with interrupted suture closure of the resulting defect is a far better alternative than four-quadrant biopsy with closure of the ulcer, in terms of procuring the biopsy material essential in all such cases to exclude gastric malignancy. Biopsy of these ulcers is essential, as 10% of perforated gastric ulcers are due to malignant disease (carcinoma or lymphomas). A case for partial gastrectomy with wide margins and beyond the perforated area can be made if the patient's condition is good. However, most would prefer ulcer excision and suture closure in the first instance and relying on histology to decide on the definitive radical therapy should the patient have gastric cancer.

Closure of the perforation is followed by aspiration of all the purulent fluid in the various compartments, including the pelvis, and irrigation with warm saline fluid (3 L). Insertion of drains is of doubtful benefit and is not generally recommended, unless difficulties are encountered with closure of the perforation. There is controversy regarding the additional definitive surgical treatment in these patients at the time of emergency surgery for closure of the perforation. In practice nowadays, this consideration arises only in those patients with a long history of peptic ulceration and in whom the perforation is early (within 6–8 hours) and whose *H. pylori* status is not known. This constitutes a rare group in modern surgical practice. If a definitive surgical treatment is considered necessary by the surgeon for any reason, this should consist of a highly selective vagotomy rather than a truncal vagotomy and drainage. Partial gastrectomy is only considered for perforated gastric ulcer situated in the body of the stomach, in which case the resection should include the ulcer. Restoration of continuity using a gastroduodenal anastomosis (Billroth I) is preferable to a Pólya gastrectomy (Billroth II) in these cases as it is less likely to cause major postgastrectomy symptoms.

Laparoscopic surgical treatment

It is important to stress that the laparoscopic approach starts as an inspection to confirm the diagnosis and to determine whether the perforation is free or sealed. In the latter eventuality, the purulent exudates are aspirated completely, and then the region of the perforation is inspected for several minutes. If there is no duodenal leakage, the natural sealing by the omentum should not be disturbed and the surgeon, following the insertion of a drain down to the sealed perforation, should proceed with warm saline irrigation with copious amounts of fluid. An effective manoeuvre during this stage is to shake the patient from side to side through the drapes to dislodge any pocketed purulent exudates from the paracolic gutters and pelvis. The management of a free perforation is the same as described with the open approach. In all other respects, the laparoscopic management is identical to that used in the open surgical approach.

Following closure of the perforated peptic ulcer, two important factors determine future management:

- the *H. pylori* status
- the requirement of the patients for NSAIDs.

All these patients require testing for *H. pylori*, and if positive eradication therapy is instituted. Some advocate eradication therapy in all patients who have sustained a perforated duodenal ulcer, but this management policy is not recommended in regions of low *H. pylori* infection prevalence. If the patient has been taking and requires continued NSAID medication, consideration should be given to a selective cyclo-oxygenase-2 inhibitor. If this proves ineffective, the addition of a PPI may reduce the future risk of recurrent ulceration.

Pyloric stenosis (gastric outlet obstruction)

Pyloric stenosis is, in fact, rarely due to stenosis at the pylorus. More commonly, the site of obstruction is on one side or other of the pylorus – either the first part of the duodenum at the site of chronic scarring from duodenal ulceration or in the antrum consequent on scarring from a benign peptic ulcer of cancer. True pyloric stenosis can arise from a pyloric channel ulcer or, very rarely, from a congenital web or adult hypertrophic pyloric stenosis. Other stenotic complications of peptic ulcer disease include hourglass and teapot deformities (gastric ulcer). All complications result in varying degrees of outlet obstruction. The stenotic complications arise from repeated cycles of ulceration and healing, with resultant dense fibrosis, deformity and narrowing. Some instances of pyloric stenosis are, however, caused by inflammatory oedema surrounding an active ulcer, and these often resolve with medical conservative treatment.

Considerable hypertrophy occurs in response to outlet obstruction. Sometimes, difficulty is encountered in distinguishing between benign gastric outlet obstruction and obstructing antral gastric cancer. This requires repeated endoscopy with biopsy, which always clarifies the exact diagnosis. In practical terms, the common causes of gastric outlet obstruction are:

- chronic duodenal ulceration with fibrosis
- antral gastric carcinoma
- carcinoma of the head of the pancreas.

Rare causes of gastric outlet obstruction

Rare causes of delayed gastric emptying include a variety of benign tumours, lymphomas, Crohn disease, duodenal haematoma (blunt trauma), adult pyloric hypertrophy, annular pancreas, mucosal diaphragm and Wilkie disease.

Adult pyloric hypertrophy

Thickening of the circular muscle of the pylorus sufficient to produce outlet obstruction can occur in adults. The relationship to congenital pyloric stenosis is unclear, although about 25% of such adult patients give a history dating back to childhood. At operation, a focal or generalized thickening of the pylorus is found. There is nearly always a degree of fibrosis so that pyloroplasty rather than pyloromyotomy is the usual corrective procedure of choice.

Mucosal diaphragm

Symptoms due to incomplete diaphragm are often not apparent until middle age. Presumably, muscular hypertrophy of the stomach muscular wall is capable of overcoming the obstruction until this time. The diaphragm consists of mucosa and submucosa and may be located in the antrum, pylorus or duodenum. It represents failure of recanalization of the embryonic foregut at this point. Gastric ulcers are sometimes found in association with this lesion. Excision of the diaphragm, with or without pyloroplasty, is all that is needed for relief of the obstruction.

Megaduodenum

Rarely, a dilated stomach is found without an organic obstruction with the dilatation extending into the duodenum to a variable extent. Such cases have been well documented as being associated with degeneration of the myenteric nerve plexus. This may be part of *Chagas disease*, but often the exact cause of the degeneration is not known. Gastrojejunostomy is usually beneficial in the short term, although progression of the degeneration to other parts of the gastrointestinal tract may occur.

Annular pancreas

Many instances of annular pancreas present in adult life with symptoms of pyloric stenosis and there is an established association of this congenital anomaly with Down syndrome and other congenital abnormalities. Annular pancreas is thought to arise because the free end of the ventral pancreas becomes fixed during embryonic development and as it migrates around to the right posteriorly to join the dorsal analogue it then encircles the duodenum (Figure 23.10). Morphologically, the condition consists of a ring of pancreatic tissue which partially or totally encircles the second part of the duodenum. The commonest presenting symptom in adult patients (usually between 20 and 40 years) is colicky upper abdominal pain (86%). Other symptoms include postprandial fullness, nausea, vomiting, weight loss and episodes of gastrointestinal bleeding. The precipitating cause may be an attack of pancreatitis. Symptomatic cases may be treated with any of the following, depending on the extent of the duodenal pathology at operation: duodenoduodenostomy, duodenojejunostomy or gastrojejunostomy.

Arteriomesenteric compression (Wilkie disease)

In this condition, the fourth part of the duodenum is compressed between the superior mesenteric vessels which cross it anteriorly such that the duodenum is compressed by the aorta behind. Acute weight loss and immobilization in a plaster cast are cited as predisposing factors to this rare condition. The symptoms are of high small bowel obstruction following a long period of insidious symptoms consisting of cramp-like epigastric pain often misdiagnosed as peptic ulcer disease. In chronic cases, megaduodenum with gastric stasis develops. However, it is important to stress that, in some patients diagnosed with *Wilkie disease*, obstruction is caused by other disease, the most important being tumours (lymphoma or carcinoma) at the duodenojejunal junction. Thus a thorough investigation is needed in all cases to establish the definitive diagnosis. Surgical treatment of Wilkie disease is either by duodenoduodenostomy (which may be difficult) or duodenojejunostomy (often easier technically) or gastrojejunostomy, which requires vagotomy or long-term acid secretory suppression.

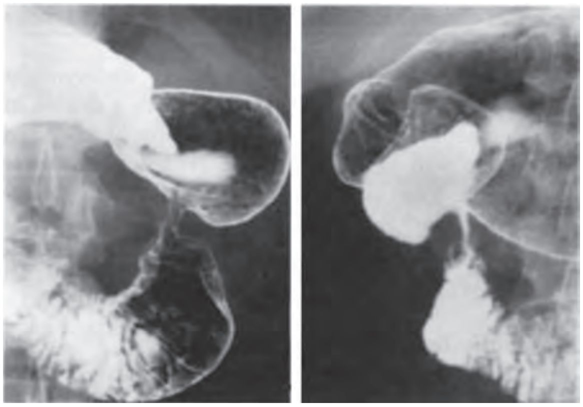


Figure 23.10 Barium meal in an adult male patient with annular pancreas who presented with intermittent episodes of vomiting after meals.

Clinical features of gastric outlet obstruction

Pyloric stenosis due to a duodenal ulcer usually occurs in a patient with longstanding symptoms of ulceration. A short preceding history with little in the way of typical ulcer pain suggests that the obstruction may be malignant. In most Western countries, the majority of gastric outlet obstruction is caused by distal gastric or proximal pancreatic cancer. In the typical case of benign pyloric stenosis, the patient experiences yet another exacerbation of the ulcer symptoms. As the obstruction develops, however, the character of the pain may change to become more cramp-like with a persistent upper abdominal discomfort. Vomiting and anorexia supervene. As vomiting increases, pain may become less of a feature. The typical vomiting of pyloric stenosis is effortless and projectile and the vomitus is characterized by an absence of bile and the presence of partially digested food eaten hours or days previously. With repeated vomiting and failure to eat, the patient often becomes constipated, although, in some cases, diarrhoea may develop.

Examination will usually show an underweight patient who is dehydrated and often with a degree of iron-deficiency anaemia. In such relatively advanced cases, there will always be evidence of gastric stasis in the form of a succussion splash. Visible peristalsis may be apparent passing across the upper abdomen from left to right, and the dilated stomach may actually be palpable.

Metabolic features

Prolonged vomiting of gastric juice and contents will result in a characteristic series of electrolyte disturbances. Initially, the major loss is fluid rich in hydrogen and chloride ions so that the dehydration is accompanied by hypochloraemic alkalosis. At this stage, the serum sodium is usually normal and the hypokalaemia may not be obvious. The more marked metabolic changes which inevitably accompany unrelieved gastric outlet obstruction result from a combination of continued losses and secondary changes in renal function. In the early stages, the urine is characterized by low chloride content and is appropriately alkaline because of enhanced bicarbonate excretion. This tends to compensate for the metabolic alkalosis, but it does so at the expense of sodium loss. If the gastric losses continue, the patient becomes progressively more dehydrated and hyponatraemic. In an attempt to conserve the circulating blood volume, sodium is retained by the kidneys and hydrogen ions and potassium are excreted in exchange. At this late stage, therefore, the patient with a metabolic alkalosis will have, paradoxically, acid urine. This further increases the severity of the alkalosis and the hypokalaemia. As a secondary effect of the alkalosis, the concentration of plasma ionized calcium may fall such that disturbances of the conscious level and tetany may be apparent.

Management

Gastric outlet obstruction due to chronic duodenal ulcer is treated by truncal vagotomy and gastroenterostomy. Other benign causes not associated with avid hypersecretion are managed with gastroenterostomy alone. Hourglass deformity and the teapot stenosis, which are complications of gastric ulcer disease, are best managed by gastrectomy if the patient's condition permits.

The priority in management of advanced cases of pyloric stenosis is correction of the fluid and electrolyte disturbances. Blood transfusion may be needed. This will often become apparent as dehydration is corrected and the spuriously high haemoglobin value falls to its actual level. Rehydration should be achieved by saline infusions with potassium supplement as indicated by electrolyte determinations. Provision of adequate sodium allows the excretion of alkaline urine so that the alkalosis is corrected. Success is indicated by clinical improvement in the state of hydration, by an increase in the urine output, a fall in the blood urea and haematocrit, and restoration to normal of the serum electrolytes.

Gastric lavage should be performed with a wide bore tube using saline for irrigation. This should be performed twice daily initially and until the returning fluid is quite clear of particulate matter. The patient should not be allowed to eat, but fluids may be given, and milky drinks or elemental diets should be encouraged. One benefit of bed rest, rehydration, lavage and milk drinks is that an ulcer will begin to heal and with subsidence of inflammatory changes, the obstruction will begin to remit especially with the administration of a PPI intravenously. Such improvement is often apparent even when the cause of obstruction is a malignancy. There must be no undue rush over this stage of management, which will often take a week to 10 days. The objective is to get the patient into the best possible condition for surgery, and it is a mistake to accept less than this ideal. During this time, the patient will also benefit from chest physiotherapy. In nutritionally compromised patients, provision of adequate nutrition often requires intravenous feeding. It may be difficult to make a firm diagnosis of the cause of the pyloric stenosis by contrast radiology. Even when the stomach has been well prepared, radiological studies may merely confirm gastric outlet obstruction and fail to reveal its cause. Flexible endoscopy with biopsy will establish the exact diagnosis and is indicated once the stomach is cleansed.

Sometimes pyloric stenosis first manifests itself when active ulcer disease within the duodenum has burnt itself out. A non-operative approach is justified in these cases, particularly in elderly or unfit patients. Balloon dilatation of the stenosed area, via an endoscope, can relieve the obstruction. Although this may need to be repeated at intervals of a few weeks for a period, this approach can obviate the need for surgery. Alternatively the endoscopic insertion of a metallic stent can be used, especially if dilation is followed by restenosis and recurrence of obstructive symptoms.

Volvulus

Gastric volvulus is rare and in terms of aetiology is classified into *type 1* (idiopathic) and *type 2*, which is either congenital or acquired. Type 1 is by far the commonest and constitutes two-thirds of the reported cases. It is thought to result from abnormal laxity of the gastrosplenic, gastrophrenic and gastrohepatic ligaments, and, indeed, cases associated with a wandering spleen have been reported. Although it usually presents in adults, cases of type 1 gastric volvulus are well documented in children. Type 2, which accounts for one-third of all cases, is associated with congenital or acquired abnormalities which result in undue mobility of the stomach anchorage, particularly at the hiatus and pylorus, such that these points of tethering become stretched and weakened. The most common cause for this in adults is hiatal hernia, particularly of the paraoesophageal type, although it may complicate all large hiatal hernias irrespective of type and indeed even diaphragmatic hernias. Rare causes of type 2 gastric volvulus include connective tissue disorders, which cause laxity of the fascial tissues (e.g. *Ehlers–Danlos syndrome*), and bulky benign gastric tumours, causing lengthening of the connective tissue attachment of the stomach.

In anatomical terms, volvulus of the stomach can occur along two axes – *mesentericoaxial* or *organoaxial* (Figure 23.11). In all types of gastric volvulus, the presentation may be acute or chronic with symptoms of recurrent episodes of epigastric pain and vomiting (Figure 23.12). Acute volvulus, more common with the organoaxial variety, presents as sudden severe pain in the epigastric or left upper quadrant region with ineffectual

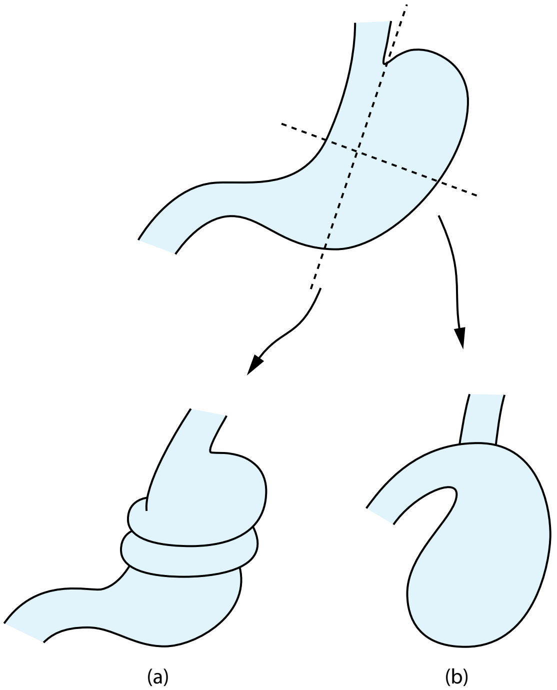


Figure 23.11 Schematic representation of the two types of gastric volvulus: (a) organoaxial; (b) mesentericoaxial.

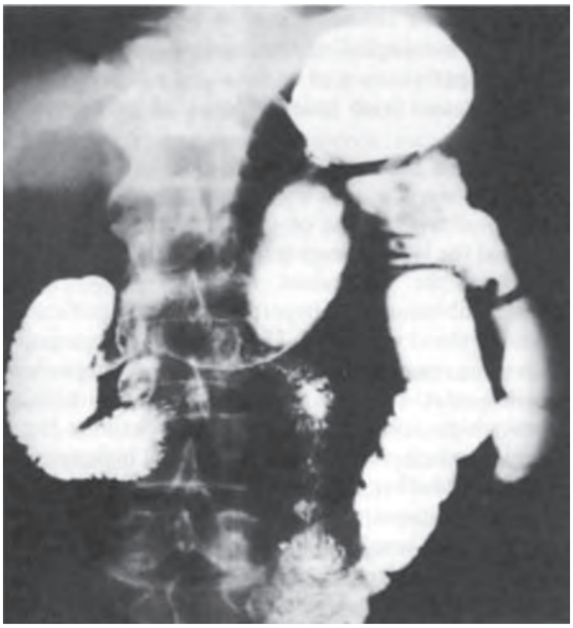


Figure 23.12 Contrast study in a patient with chronic mesentericoaxial volvulus.

retching. When the volvulus is intrathoracic, the pain is located in the chest and radiates to the neck, shoulder, arms and back, simulating acute myocardial infarction. A few patients may present with haematemesis secondary to gastric wall ischaemia. Distension, upper abdominal tenderness and signs of shock develop rapidly in these acute cases, and urgent surgical intervention is indicated if strangulation and perforation are to be avoided. With this clinical background, the inability to pass a nasogastric tube into the stomach is diagnostic in 70% of patients.

Endoscopic reduction should be attempted but only in the absence of clinical signs of strangulation. The flexible endoscope is negotiated through the twist into the stomach, where deflation and untwisting occur. If successful, the patient is managed conservatively, especially if the general condition is poor and in the presence of comorbid disease. Otherwise urgent surgery is indicated. At operation, the volvulus is untwisted to restore tissue perfusion and, if the stomach walls are viable, the stomach is fixed by a gastropexy with the insertion of a percutaneous endoscopic gastrostomy tube. Partial or complete resection may be necessary in cases of delayed diagnosis with established gastric gangrene.

Haematemesis and melaena

Patients presenting with acute upper gastrointestinal haemorrhage still pose management problems, and despite the advances in medical treatment, including diagnostic and interventional endoscopy, the overall mortality from acute upper gastrointestinal bleeding has not changed appreciably during the past two decades and probably averages 20%, although there is sufficient evidence from good prospective and retrospective reports that, with the appropriate management strategy, the hospital mortality of these patients can be reduced to 5%.

Aetiology

The causes of upper gastrointestinal haemorrhage are discussed in Chapter 5. The vast majority (>90%) are caused by chronic peptic ulceration, NSAID-induced bleeding and oesophagogastric varices. On occasions the source of an upper gastrointestinal bleed may remain unknown. If the presenting problem is melaena alone and no blood is found in the stomach or duodenum on repeat endoscopy, then attention should turn to the small bowel and the colon, and certainly capsule small bowel endoscopy is indicated as this in now the investigative procedure of choice and has replaced all other tests for the diagnosis of obscure bleeding from the small intestine. If small bowel capsule endoscopy proves non-productive, emergency mesenteric angiography is indicated. In any case, there is no indication for emergency laparotomy for undiagnosed gastrointestinal bleeding in modern surgical practice.

Chronic peptic ulcers

Despite more effective medical treatment, bleeding from duodenal and gastric ulcers still remains one of the common causes of life-threatening upper gastrointestinal haemorrhage. Some consider the bleeding caused by NSAIDs to be largely due to exacerbation of chronic peptic ulcer disease, especially gastric ulcers in elderly women. The problem inherent to bleeding peptic ulcers is their tendency to rebleed after spontaneous or therapeutic arrest. This propensity to recurrent haemorrhage has been attributed to the acid environment of the stomach and the proximal duodenum that impairs platelet aggregation and blood coagulation and to the digestion of clots by pepsin. In an interesting study on human volunteers, intraduodenal infusion of blood resulted in decreased pentagastrin-stimulated acid and pepsin secretion, suggesting a natural protective mechanism.

NSAID-induced bleeding

Aspirin and non-aspirin NSAID usage in elderly people is widespread in the general population, and population-based studies document annual prevalence rates as high as 60%. A meta-analysis of 16 studies has shown that NSAID users have a threefold risk of gastrointestinal haemorrhage, surgery and death compared with non-users. The risk from bleeding is greatest:

- in the first few months of treatment
- in the elderly (> 65 years)
- in patients with concomitant steroid use
- in patients with a previous history of gastrointestinal events.

The incidence of bleeding and perforation according to the Nottingham study is 1:6000 and 133 000 prescriptions respectively. Of all the NSAIDs known to cause bleeding or perforation (indometacin, naproxen, etc.), aspirin produces the most damage. There is some evidence that the newer NSAIDs, e.g. nabumetone, which are selective inhibitors of cyclo-oxygenase-2, are less damaging to the gastroduodenal mucosa and hence significantly less ulcerogenic, but they appear to be less effective in relieving pain. The other problem with NSAIDs concerns the development of non-specific ulceration of the upper small intestinal mucosa in 8–9% of users. These can bleed and perforate.

Dieulafoy's lesion (exulceratio simplex, cirsoid aneurysm)

This condition was first described in 1898 by a French surgeon. It consists of a nodule containing an arteriole which protrudes through a mucosal defect and is typically located within 6 cm of the oesophagogastric junction on the lesser curvature. Extragastric Dieulafoy's lesion may occur in other parts of the gastrointestinal tract: oesophagus, duodenum (usually periampullary), small bowel, colon and rectum. The condition is nowadays thought to be caused by an abnormally large-calibre tortuous submucosal artery which erodes the overlying gastric mucosa by the repeated incessant arterial pulsations.

Dieulafoy's lesion is an uncommon cause of recurrent gastrointestinal bleeding and accounts for 1% of all cases of acute, often massive, upper and lower non-variceal gastrointestinal bleeding. The diagnosis is difficult and is usually made after repeated endoscopies. The most common presentation is with recurrent, often massive, haematemesis associated with melaena (51%) or haematemesis alone (28%) and, least commonly, melaena alone (18%) without any accompanying dyspeptic symptoms.

Therapeutic endoscopy is the treatment of choice with electrocoagulation, sclerotherapy, heater probe, laser photocoagulation, epinephrine injection, haemoclipping and banding, depending on the local availability and expertise. Endoscopic therapy is successful in achieving permanent haemostasis in the majority (85–90%) of cases. Rebleeding after treatment is initially treated by further endoscopic therapy but some 5% ultimately require surgical treatment (gastrotomy with suture ligation). Bleeding from extragastric Dieulafoy lesions can be treated by angiographic therapeutic Gelfoam embolization.

Portal hypertensive gastropathy

This condition can cause acute or chronic repeated gastrointestinal blood loss. Up to 65% of patients with portal hypertension from cirrhosis will develop portal hypertensive gastropathy (PHG) but equally it occurs in patients with non-cirrhotic portal hypertension (presinusoidal). PHG is usually associated with the presence of oesophageal and/or gastric varices. The pathogenesis of PHG is not fully understood but regulation of gastric nitric oxide, prostaglandins, tumour necrosis factor (TNF)- and EGF production may be involved. The reported data on seroprevalence of *H. pylori* and PHG are conflicting and currently there is no clear correlation between *H. pylori* infection and aetiology, stage of cirrhosis and presence/severity of PHG. Additionally, there is no clear correlation between PHG and the severity of the cirrhosis and degree of portal hypertension. The histological appearances of PHG consist of mucosal and submucosal vascular dilatation with little evidence of inflammatory changes.

The diagnosis of PHG is made by endoscopy. In mild PHG, the gastric mucosa often looks reddened and oedematous with a faintly mosaic pattern. With established disease, the endoscopic appearances predominantly affect the fundus, but no region is exempt and the condition may be generalized. The mucosa often has a marked mosaic-like pattern (Figure 23.13) and may exhibit red point lesions, cherry red spots, and black-brown spots. Treatment of PHG is by beta-blockers, which reduce the frequency of bleeding episodes. Removal of all risk factors known to cause gastric injury, e.g. alcohol, aspirin and NSAIDs, is important for success of medical therapy. If this fails, surgical decompression of the portal system, or nowadays by radiologically guided transjugular intrahepatic stenting and liver transplantation (in cirrhotic patients), will effectively resolve this condition.

Gastric antral vascular ectasia or watermelon stomach

This is a related disorder but must be distinguished from PHG. The exact pathogenesis of gastric antral vascular ectasia (GAVE) syndrome remains unclear, although one hypothesis attributes the cause to repeated trauma and ischaemia from antroduodenal prolapse. The classical endoscopic features of GAVE syndrome consist of hyperaemic antral folds with dilated, tortuous vessels extending radially from the pylorus, producing a watermelon appearance accounting for the alternative name. Ectatic submucosal capillaries, microvascular fibrin thrombosis and fibromuscular hyperplasia in the lamina propria constitute the distinctive histological features. Although the majority (70%) of patients with GAVE syndrome do not have cirrhosis or portal hypertension, cases associated with cirrhotic and non-cirrhotic portal hypertension do occur and must be differentiated from PHG; whereas PHG generally responds to a reduction in portal pressures, GAVE associated with portal hypertension does not. Cases of GAVE associated with other disorders including end-stage chronic renal disease are documented

GAVE presents with upper gastrointestinal haemorrhage. The bleeding is often recurrent, requiring multiple transfusions. Treatment options include endoscopic therapy and surgery, which is performed only if endoscopic therapy fails. Flexible endoscopic therapy (electrocoagulation, sclerotherapy, laser coagulation) constitutes the first line of management and is effective in the majority, although it often has to be repeated.

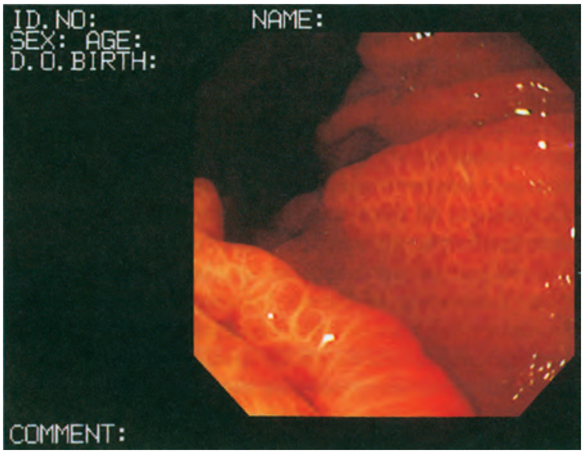


Figure 23.13 Mosaic-like appearance of portal hypertensive gastropathy. (Courtesy of Professor John F Dillon, Dundee, UK.)

Surgical treatment consists of distal gastrectomy (open or laparoscopic) and is always curative but accompanied by a reported high postoperative mortality.

Tumours

Gastrointestinal haemorrhage may be caused by both benign and malignant tumours. Acute haemorrhage is, however, more commonly associated with benign lesions, such as neurofibromatosis, smooth muscle tumours and gastrointestinal stromal tumours (GISTs). Malignant tumours (carcinoma and lymphomas) more usually cause chronic blood loss with the development of iron-deficiency anaemia, although massive bleeding may be precipitated by combination chemotherapy.

Chemotherapy

Life-threatening bleeding or perforation from necrosis of the tumour may complicate chemotherapy for gastrointestinal tumours, especially lymphomas. The bleeding is aggravated by the frequent thrombocytopenia induced by the treatment. In addition to haematological support, surgical intervention with resection of the tumour (if possible) is indicated to save life.

Stress ulceration

There has been a substantial decrease in the incidence of stress ulceration in intensive care units and this has been attributed to improved methods of supportive care. Stress ulceration causing acute bleeding is treated conservatively in the first instance by intravenous PPI and sucralfate together with blood transfusion. Sucralfate is administered via a nasogastric tube after aspiration to remove blood in aliquots of 60 mL repeated every 2–4 hours. Surgical treatment is only considered if conservative treatment fails, which is rare.

Aortoenteric fistula

This usually complicates aortic replacement with prosthetic grafts but may arise spontaneously. The aetiology is thought to be the result of graft infection. Small repeated warning haemorrhages usually precede the catastrophic bleeding, which when it occurs often leads to fatal exsanguinations, although prompt aggressive surgical intervention with ligature of the aorta, removal of the graft and axillofemoral bypass may save some of these patients.

Duodenal diverticulum

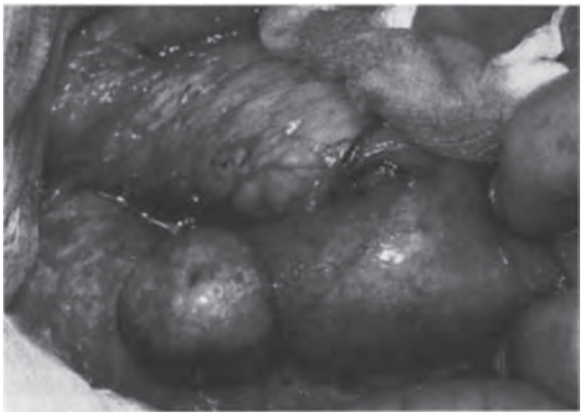


Figure 23.14 Duodenal diverticulum in a patient who failed to respond to conservative treatment and required open surgical intervention.

Although fairly common (Figure 23.14) and usually located in the ampullary region, bleeding from a duodenal diverticulum is rare (1% of cases). Acute haemorrhage may be caused by erosion of a major vessel, bleeding ectopic gastric mucosa, an intradiverticular polyp or local inflammatory process. Duodenal diverticula are managed conservatively unless complications (bleeding or perforation) arise. Aside from causing difficulties with cannulation of the bile/pancreatic duct during endogastric retrograde cholangopancreatography, a solitary duodenal diverticulum increases the risk of duodenal perforation in patients undergoing endoscopic sphincterotomy. The treatment of a bleeding duodenal diverticulum is conservative (endoscopic sclerotherapy or clip ligation) in the first instance, but surgical treatment may be required if endoscopic treatment fails or the bleeding recurs after successful control. At operation the procedure involves duodenotomy of the second part with eversion of the diverticulum (by sutures) and suture ligation of the bleeding vessel.

Management

The successful treatment of acute upper gastrointestinal haemorrhage demands aggressive treatment and a policy that is flexible enough to deal with the problems peculiar to the individual patient. There is good evidence that the best outcome is achieved by combined management between gastroenterologists and surgeons working as a team and in close co-operation from the time of admission of the patient. There are three phases in the management of the bleeding patient: resuscitation, diagnosis and definitive treatment (see Chapter 5).

Acute dilatation of the stomach

Acute dilatation of the stomach is a serious condition which may cause death by aspiration. Although encountered most commonly as a complication of upper abdominal surgery (especially splenectomy) and pelvic surgery, it occurs in other situations including orthopaedic patients immobilized in plaster casts. It is more common in malnourished and debilitated patients. Other predisposing factors include aerophagy (by apprehensive patients), excessive inadvertent distension of the stomach during endotracheal intubation and ventilation, administration of oxygen by nasal catheter, flexible endoscopy with vigorous insufflations of gas and the use of opiate analgesia. Contrary to expectations, acute dilatation of the stomach is rarely encountered in gastric outlet obstruction. The exact aetiology is unknown but is likely to be multifactorial. The stomach becomes atonic and this has been attributed to reflex inhibition of the myenteric neurones supplying the gastric musculature or failure of the gastric pacemaker.

Acute dilatation of the stomach is well documented in patients (usually young females) suffering from *anorexia nervosa* or *bulimia* (compulsive eating followed by self-induced vomiting), and instances of gastric necrosis and rupture of the stomach have been reported in these patients. The suggested mechanisms for acute dilatation in anorexia nervosa include muscular atrophy from starvation and neurogenic paralysis. In both conditions gastric necrosis supervenes when the intragastric pressure exceeds the gastric venous pressure.

The hugely dilated stomach, which may occupy the whole of the abdomen, is filled with dark blood-stained fluid. Evidence of hypovolaemia due to fluid and electrolyte sequestration is often present and the patient is frequently hypokalaemic. Premonitory symptoms include hiccups, vague feelings of unease in the epigastric region and small vomits which contain altered blood. A gastric succussion splash is diagnostic. The more dramatic presentation, which usually follows these signs and symptoms but which may arise *de novo,* is either severe pain mimicking myocardial infarction or severe collapse from hypovolaemia (simulating pulmonary embolism) or sudden marked vomiting of large amounts of foul-smelling fluid with inevitable aspiration and the development of acute respiratory distress syndrome – Mendelson syndrome.

The treatment of this emergency situation is by prompt decompression of the stomach by a large-bore nasogastric tube, preferably of the Salem sump suction variety, and correction of the hypovolaemia and electrolyte deficit by crystalloid solutions. Pulmonary aspiration is treated by bronchoscopic suction and lavage, antibiotics and steroids together with endotracheal intubation and ventilatory support in the intensive care unit. The mortality of acute gastric dilatation accompanied by pulmonary aspiration remains high. Early detection of the condition with appropriate intervention will prevent the vast majority of these deaths.

Nasogastric suction

Routine nasogastric suction by a Ryle's tube was introduced as a precaution against complications such as acute gastric dilatation in all patients undergoing abdominal surgery. This practice is now no longer recommended as a number of prospective controlled clinical trials have demonstrated (level 1 clinical evidence) that the routine use of nasogastric suction after elective surgery is accompanied by a higher incidence of postoperative infective complications, delays recovery, increases hospital stay, aside from causing considerable discomfort to the patient.

Although routine prophylactic nasogastric decompression has been abandoned in modern surgical practice, early therapeutic nasogastric suction is mandatory in all patients in whom the postoperative ileus persists beyond 24–48 hours after surgery. It is also indicated in acute abdominal conditions, e.g. intestinal obstruction, perforation, gastrointestinal bleeding, peritonitis, acute gastric dilatations, etc. Therapeutic nasogastric suction is best achieved by the use of the Salem sump suction tube rather than the cheaper but more popular single-lumen Ryle's tube. If a nasogastric tube is inserted, then continuous drainage with intermittent aspiration to establish continued patency is essential. There is never an indication to plug the tube, as this amounts to bad practice since, aside from causing patient discomfort, the tube is not contributing any possible benefit to the patient under these circumstances.

Peptic ulcer disease

Although there are differences in the clinical features between the two, it is customary to consider gastric and duodenal ulcers together as peptic ulcer disease.

Epidemiology

In the Western world, the incidence and prevalence of peptic ulcer disease and its complications increased from the beginning of the twentieth century to reach a peak in the period 1960–1970 and thereafter declined. During the same period, the incidence of peptic ulcer disease (especially duodenal ulcers) started to increase in other parts of the world, such as Kenya and South Africa, suggesting an increased exposure to environmental ulcerogens. Worldwide, duodenal ulcers are more common than gastric ulcers and there is a significantly higher incidence of duodenal ulcers in males of all age groups. In Africa and India, almost all ulcers are duodenal, and stenosis with obstruction is a frequent complication. In Europe, duodenal ulcers are two to four times as common as gastric ulcers but there are some regional variations, e.g. duodenal ulcers are twice as frequent in Scotland as in England. Duodenal ulceration is common in North America but gastric ulcer appears to be less common than in Europe. Although duodenal ulcers are prevalent in Australia, a relatively high incidence of gastric ulcers is encountered in younger females.

Aetiology

Aside from genetic factors (increased prevalence of duodenal ulcers in patients with blood group O who are non-secretors and in individuals with hyperpepsinogenaemia 1), dietary factors, drug ingestion (NSAIDs) and smoking are important. The last is associated both with an increased incidence of gastric and duodenal ulceration and with a higher relapse rate following successful healing. The most important cause is however infection with *H. pylori*, which is acquired in childhood. This pathogen is responsible for peptic ulceration (duodenal and gastric) though not all infected individuals develop ulcers. Duodenal ulcer, however, develops in all patients with antral predominant infection and who also have a large parietal cell mass, thus augmenting the hypersecretion of HCl. The risk of peptic ulceration is determined also by the severity of the gastritis. The vacuolizing toxin secreted by *H. pylori* is involved in the development of duodenal ulceration. Undoubtedly the infection by impairing the mucus–bicarbonate protective layer plays an important role in the chronicity of both duodenal and gastric ulcers and their tendency to relapse. The only permanent healing is achieved by eradication of the *H. pylori* infection. Strains of *H. pylori* with vacA signal sequence type S1A are associated with severe gastritis and duodenal ulcers, whereas vacA S2 strains cause a milder gastritis with little tendency to ulceration. There is also evidence that strains of *H. pylori* that produce cytotoxin and induce a neutrophil oxidative burst are associated with peptic ulcer disease.

Chemical ulcerogens include substances in food and drugs. Information on dietary ulcerogens is scanty apart from the association with excessive alcohol, smoking and coffee and Coca-Cola consumption in the West. Although chronic ingestion of hot spicy food is often incriminated, there is little epidemiological evidence to support this. By contrast there is some evidence on dietary protective factors, although the exact mechanism involved in not known. Dietary factors include high-fibre diets, increased consumption of essential fatty acids (-linolenic acid, which is a prostaglandin precursor) and fruit and green vegetables (particularly raw cabbage, Brussels sprouts and lentils).

The most important group of chemical ulcerogens is constituted by aspirin and other NSAIDs and steroids. They are the commonest cause of peptic ulceration in *H. pylori*-negative individuals. However, these drugs cannot be regarded as specific gastroduodenal ulcerogens since they also induce damage and ulceration of the small and large intestine. Other drugs that can cause both gastric and duodenal ulcers include cocaine and amphetamine (usually drug addicts). There are some important differences between ulcers caused by *H. pylori* and those caused by NSAIDs. These include:

- NSAID-associated ulcers are more likely to cause gastrointestinal haemorrhage; thus overall, 75% of patients with upper gastrointestinal bleeding from peptic ulcers are on NSAID medication
- gastric ulcers caused by *H. pylori* are rarely encountered on the greater curve (5%), being most commonly situated on the lesser curve (85%), whereas NSAID ulcers (in the absence of *H. pylori* infection) occur along the lesser and greater curvatures in 35% and 45% of patients, respectively.

H. pylori infection and NSAID usage is encountered in 20% of patients. Eradication of the infection does not appear to influence the healing and recurrence of peptic ulcers associated with chronic NSAID medication.

Role of acid and pepsin

The maxim 'no acid, no ulcer' is no longer tenable. Although 30–40% of duodenal ulcer patients exhibit acid hypersecretion and there are well-established syndromes of persistent gastric hypersecretion associated with intractable ulceration, the overlap between the acid secretory status of duodenal ulcer patients and controls is considerable and there are many normal individuals who hypersecrete acid (with large parietal cell mass) who do not develop ulcers in the absence of infection by *H. pylori*. Furthermore duodenal ulcers can develop in individuals with normal gastric acid output and almost all patients with gastric ulcers have normal or reduced acid secretion. Nonetheless, gastric acid is an important factor in the chronicity of the disease and acid suppression with antibiotic therapy is necessary for *H. pylori*-associated peptic ulcer disease.

The secretory characteristics of the usual duodenal ulcer patients (the norm) include an increased acid secretory capacity due to an enlarged parietal cell mass and enhanced maximal acid output in response to pentagastrin, increased gastrin response to food (protein) and insulin, increased sensitivity to gastrin and defective inhibition of acid secretion normally elicited by antral acidification, antral distension and intraduodenal fat. There is an increased concentration of pepsins in the gastric juice of patients with duodenal ulceration, especially pepsin I, which is the more mucolytic. The disruption of the mucus–bicarbonate layer by pepsin I exposes the underlying mucosa to injury by ulcerogens and impairs healing by removal of the protective mucous cap (blister effect).

Other factors

Enterogastric reflux of bile salts and lysolecithin with destruction of the mucus–bicarbonate layer and mucosal injury was first suggested by Capper as the cause of gastric ulceration. Although bile salts in particular can lead to back diffusion of acid following disruption of the mucous gel, this theory has never been conclusively proven. Similarly, antral stasis with delayed gastric emptying resulting in antral distension and increased gastrin release has been implicated in the development of pyloric channel and prepyloric ulcers. As a group, these gastric ulcers behave like duodenal ulcers and tend to be associated with hyperacidity as distinct from the more usual proximal gastric ulcers, which originate on a background of normal or hypoacidity and are accompanied by atrophic gastritis. Despite the extensive literature on the potential vascular cause of peptic ulcers as exemplified by instances of gastric ulceration after highly selective vagotomy and in association with mesenteric ischaemia, or experimental restriction of the blood supply to the stomach in animals, there has been no confirmation that focal vascular insufficiency is important in the pathogenesis of peptic ulcer disease.

Peptic ulcers occur at specific localized sites: antropyloric junction on the lesser curve, etc.; this has been interpreted as indicative of trauma resulting from the action of ulcerogens being specifically directed to these sites. Examples illustrative of this effect include the lesser curve location of gastric ulcers following reflux of duodenal contents and the occurrence of 'sump' ulcers following ingestion of ulcerogenic tablets which gravitate to the most dependent part of the stomach. The same mechanism has been proposed by Kirk for explaining the location of duodenal ulcers, i.e. the jets of gastric contents repeatedly hit the duodenal bulb in the same area causing focal ulceration.

Stress induces gastric hypersecretion and can lead to acute (stress) ulceration in patients. The stress may be purely psychological or result from major complicated disease (usually in intensive care patients) and after major burns (Curling's ulcer). Although psychological stress is usually an aggravating factor, it may be the primary cause in some patients. Rarely, duodenal ulcerations may be associated with other disorders, which include liver disease (particularly after shunt surgery for portal hypertension), persistent hypercalcaemia, renal failure and massive small bowel resection.

Clinical features

Chronic duodenal ulceration

This can occur at all age groups but the peak incidence is between 25 and 50 years. Whereas mortality caused by duodenal ulceration is low to negligible in this age range, the age-specific death rate (from complications of the disease) rises to 24 per

100 000 and 7 per 100 000 above 65 years in males and females respectively. This difference in mortality is due to the significantly higher incidence of duodenal ulceration in males with a male–female ratio of 2.0–4.0:1.0. Duodenal ulceration is a remitting disease characterized by periods of activity and quiescence. Exacerbations may be associated with periods of stress, dietary or alcoholic indiscretions and smoking. Early in the history of the disease, remissions may be associated with complete healing; but as the disease progresses, there is a tendency towards fibrous scarring so that evidence of past disease may be found on investigation even when the patient is free of symptoms. Typically, the epigastric pain is experienced during fasting or before meals (hunger pain) when the stomach is empty and there is nothing to buffer the acid secretion. Relief usually follows eating, ingestion of milk or alkali. Failure to produce relief, particularly if the pain radiates to the back, is suggestive of posterior ulcer penetration into the pancreas. The postprandial pain relief lasts for varying periods but usually averages several hours before it recurs and often occurs at night waking the patient up. Vomiting is not usually a feature of uncomplicated disease but may develop in severe exacerbation of an inflamed ulcer with surrounding gross oedema of the duodenal bulb or result from fibrosis causing organ outlet obstruction (pyloric/duodenal stenosis).

As the patients are constantly nibbling food to ward off painful indigestion, they are usually overweight. Other than this, the physical signs may amount to no more than diffuse epigastric tenderness, although this is sometimes well localized. Occult bleeding may produce iron-deficiency anaemia. The presence of a succussion splash indicates delayed gastric emptying.

Chronic gastric ulcer

Chronic ulceration of the stomach is less common than in the duodenum in most countries, the ratio of gastric to duodenal ulcers varying from 1:4 to 1:20. There are two quite distinct types of gastric ulcer: type I, which occurs in the body of the stomach along the lesser curve, and type II, which is referred to as the pyloric channel ulcer and includes prepyloric ulcers. The natural history, acid secretory profile and therapeutic response of type II ulcers are akin to those of duodenal ulceration. Type I chronic gastric ulcer may arise in a normal mucosa or on a mucosal background of atrophic gastritis. The disease is not associated with hyperacidity and, indeed, hypoacidity is frequently encountered, particularly in patients with atrophic gastritis. There is a male preponderance but this is not as marked as in duodenal ulceration and the peak incidence is encountered after middle age. The age-specific mortality from complicated gastric ulcer disease in patients above the age of 65 years is 57/100 000 and 41/100 000 in males and females respectively.

Gastric ulceration is most commonly encountered in late middle-aged and elderly patients. In Australia, an association with chronic ingestion of NSAIDs and the relatively high prevalence in younger females has been documented. The association with chronic ill health and lower socioeconomic class is not as obvious nowadays as it was in the past. The main clinical feature is pain, or, perhaps more commonly, a feeling of acute discomfort and fullness in the epigastrium. Unlike duodenal ulcer, the pain of gastric ulceration is not experienced during fasting when the stomach is empty. Indeed, the converse is true, as eating produces or exacerbates the pain. For this reason, patients suffering from the disease are afraid of eating and because of the reduced dietary intake they are usually underweight. It is important to stress that this symptom complex of indigestion immediately after meals and weight loss is indistinguishable from that produced by gastric cancer, and clinically it is impossible to differentiate between the two disorders unless the cancer is advanced and incurable. Nausea and vomiting are more common symptoms than in duodenal ulcers, even in the absence of outlet obstruction. Although periodicity with remissions and relapses is encountered, this is not as obvious as in duodenal ulceration.

Aside from weight loss, physical examination is not usually rewarding, with epigastric tenderness being the only fairly constant finding. Gastric ulcers may obstruct, perforate or bleed. Occasionally, a large inflammatory mass around an ulcer may be palpable, particularly if the patient has lost weight, but for all practical purposes, a palpable epigastric mass in a patient with this dyspeptic background should always indicate the probability of gastric cancer being the cause of the patient's symptoms.

Investigations

In most centres, upper gastrointestinal endoscopy has replaced double contrast barium meals because of the greater diagnostic accuracy. In patients with duodenal ulcer, the findings vary from a definite crater in the first part of the duodenum to severe duodenitis. Endoscopic biopsy of the antrum is necessary in all cases for establishing the *H. pylori* status, but the ulcer itself should not be biopsied as it is always benign and the biopsy may precipitate bleeding. Although double contrast barium meal carries a high diagnostic yield for gastric ulcer, and, if performed expertly, will differentiate benign gastric ulcer disease from early ulcerating cancer, endoscopy and multiple biopsies with brush cytology are mandatory in all patients with gastric ulceration. The *H. pylori* status is best determined on the endoscopic biopsies (histology). Furthermore, the endoscopy should be repeated after 2–3 months and if ulcer healing is not documented further biopsies are taken and surgical treatment is indicated even if the second biopsies are reported as benign.

Medical treatment of peptic ulcer disease

Modern evidence-based treatment of peptic ulcer disease is based on the aetiology of the disease in the individual patient in accordance with the following categories:

- *H. pylori*-positive duodenal or gastric ulcers
- *H. pylori*-negative duodenal or gastric ulcers
- ulcers caused by NSAID medication which may be *H. pylori* positive or negative.

Hence, the importance of testing for *H. pylori* infection in all patients. The National Institute for Health and Clinical Excellence (NICE), Guideline no. 17 (www.nice.org.uk/CG017) recommends use of the ^{13}C-urea breath test or a stool antigen test over serological tests, which have lower sensitivity and specificity (92% and 83%, respectively) than either breath testing (sensitivity 95%, specificity 96%) or the stool antigen test (sensitivity 95%, specificity 94%). The lower predictive value of serological tests (64% vs 88% for the ^{13}C-urea breath test and 84% for the stool antigen test) indicates an increased probability of false positives with serological testing such that a number of patients would be treated unnecessarily with antibiotics for eradication of the infection. The recommendation from NICE is that retesting after eradication treatment to confirm clearance of the infection should currently be performed by the ^{13}C-urea breath test as currently there is insufficient evidence to recommend the stool antigen test as a test of eradication.

Eradication therapy *for Helicobacter pylori*

All patients who test positive should receive a 7 day twice-daily course of eradication treatment consisting of a full-dose PPI, with either metronidazole (400 mg) + clarithromycin (250 mg) or amoxicillin (1 g) + clarithromycin (500 mg). This regimen is effective in 80–85% of patients. Other regimens such as PAC500 (PPI, amoxicillin, clarithromycin 500 mg) and PMC250 (PPI, metronidazole, clarithromycin 250 mg) achieve the same eradication rate.

However, PMC250 when used as a first-line therapy may induce resistance to both clarithromycin and metronidazole. Although 14 day courses have been reported to increase the eradication rate by 10% (to 95%) they do not appear to be cost effective according to NICE. In patients who require a second course of eradication therapy the antibiotic combination with the full dose PPI should not include any of the antibiotics used in the initial eradication attempt.

In patients using NSAIDs with peptic ulcer, *H. pylori* eradication does not increase healing compared with acid suppression therapy in patients using NSAIDs with a peptic ulcer. However, in patients on chronic NSAID medication without peptic ulcer disease, *H. pylori* eradication reduces the risk of a first occurrence of peptic ulcer. Furthermore in patients on NSAIDs with a history of previous peptic ulcer, *H. pylori* eradication reduces the risk of recurrence.

There is a core of peptic ulcer patients (10–15%) who are resistant to the normal eradication regimens (usually with infections by cagA strains). In these patients an alternative regime is used: PPI with triple therapy with tripotassium citratobismuthate (De-Nol tab) 120 mg q.d.s., metronidazole 400 mg t.d.s. and tetracycline 500 mg t.d.s. for 2 weeks.

Duodenal ulcer

Aside from general dietary measures, such as abstinence from alcohol, cessation of smoking and avoidance of hot spicy food, excessive coffee drinking and chocolate, management is based on PPI therapy, testing for *H. pylori* and cessation of NSAIDs. Whenever possible, NSAIDs should be stopped. Otherwise, they are changed to a cyclo-oxygenase (Cox)-2 inhibitor or paracetamol or low-dose ibuprofen. Initial treatment is with a full-dose PPI for 2 months followed immediately by a 1 week course of eradication therapy in *H. pylori*-positive patients. The PPI therapy induces healing of the duodenal ulcer in 70%, and although this is only increased by a further 5% by eradication, this dramatically reduces the ulcer recurrence rate. Thus studies have shown that at 3–12 months only 39% of patients treated with PPI alone are ulcer free, but this is increased by a further 52% by successful eradication of the infection.

Frequent relapses or failure to control the disease by on demand PPI therapy in *H. pylori*-negative patients should raise the possibility of rare causes of duodenal ulcers: gastrinoma, Crohn disease, etc.

Gastric ulcer

The medical management of gastric ulcer is similar. If the patient is on NSAIDs these should be stopped if possible or substituted as previously outlined. A full-dose PPI course is administered for 2 months and is followed by eradication treatment in *H. pylori*-positive patients. All patients with gastric ulcer require an endoscopy and retesting for *H. pylori* 6–8 weeks after the start of treatment. The main reason for the endoscopy is to establish healing or otherwise of the ulcer. If this has healed and the patient is *H. pylori* negative, low-dose maintenance treatment is individualized (low-dose PPI continuous or on demand). Patients who remain *H. pylori* positive after eradication need an endoscopy with biopsy. In patients in whom eradication is successful but the ulcer has not healed, PPI in high doses is continued for a further 4 weeks and then the patient is re-endoscoped. If the ulcer has not healed by then, it should be considered as malignant, biopsied and referred for surgical opinion.

It should be noted that *H. pylori* eradication therapy does not increase gastric ulcer healing in *H. pylori*-positive patients, when compared with acid suppression alone, but it reduces gastric ulcer recurrence. Thus studies have shown that, after 3–12 months, 45% of patients receiving acid suppression therapy alone are ulcer free. Eradication increases this by a further 32%.

Indications for surgical treatment of peptic ulcer disease

Elective surgery for duodenal ulcer disease has declined substantially during the past 30 years and is now largely restricted to the treatment of complicated or, even more rarely, uncomplicated disease that proves resistant to medical treatment (intractable disease). In this respect, the situation is changing in Western countries with the marked decline in the prevalence of *H. pylori* infection in the general population to around 8–10%. Furthermore, failure of eradication of *H. pylori* is an increasing problem and reliance on long-term acid suppression with PPIs has its problems and is costly. Laparoscopic elective peptic ulcer surgery offers a cost-effective alternative to long-term medication and is increasingly preferred to open surgery for both complicated and intractable duodenal ulcer disease.

Operations for duodenal ulcer

The aim of all operations for uncomplicated duodenal ulceration is to reduce acid secretion to such levels that the ulcer will heal permanently. In the majority of patients, the surgeon is faced with the choice of either gastric resection (to reduce the size of the parietal cell mass and remove the antrum and thus gastrin release) or vagal denervation of the stomach (to abolish the cephalic phase of secretion and reduce the sensitivity of the parietal cell mass to secretory stimuli), or a combination of the two. Although the purpose of the operation is to achieve permanent cure of the ulcer diathesis, this should not be the only consideration, as other factors are important in the decision-making process regarding the selection of the correct operation in the individual patient: operative morbidity and mortality, and the incidence of side effects need to be weighed carefully against success rate in terms of healing and ulcer recurrence rates.

The relative advantages and disadvantages of gastrectomy, vagotomy and drainage and highly selective vagotomy are summarized in Figure 23.15. Subtotal gastrectomy with gastrojejunal anastomosis is an excellent antiulcer operation, but its merits are seriously marred by the increased morbidity/mortality as well as the high incidence of troublesome side effects and nutritional sequelae. The initial hope that the operation of truncal vagotomy and drainage would avoid the alimentary sequelae and maintain good nutrition did not materialize and some surgeons are dissatisfied with this procedure. Truncal vagotomy and antrectomy possesses the advantage of a dual attack on the acid secretory mechanisms and, as a result, carries the lowest recurrent ulcer risk. Unfortunately, it incurs the disadvantages of both resection and vagotomy. Truncal vagotomy and antrectomy is probably the operation of choice for intractable and resistant ulcers including pyloric channel and prepyloric ulcers, as these are attended by high ulcer recurrence rates.

Vagotomy procedures

The alternatives to truncal vagotomy and drainage (by gastrojejunostomy or pyloroplasty) are bilateral selective vagotomy with drainage, highly selective vagotomy (parietal cell, proximal gastric, selective proximal gastric vagotomy), posterior truncal vagotomy and anterior seromyotomy (Taylor procedure), and posterior truncal vagotomy and anterior highly selective vagotomy (Hill procedure).

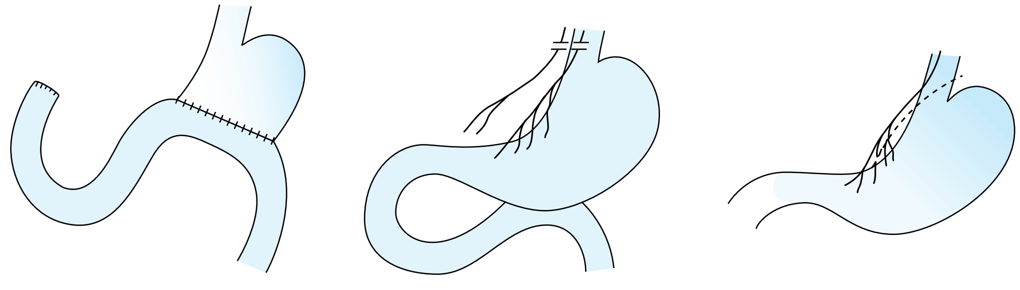
Bilateral selective vagotomy confines the denervation to the stomach and preserves the vagal branches to the liver (hepatic plexus). The operation requires additional gastric drainage as the antrum is denervated. Gastric stasis and ulceration have been documented in patients in whom drainage had been omitted. Although bilateral selective vagotomy does reduce the incidence of diarrhoea and achieves adequate gastric denervation, the overall clinical results obtained by this operation have been no better than those of truncal vagotomy with drainage, and, for this reason, the operation has been largely abandoned.

Undoubtedly highly selective vagotomy (HSV) is the most physiological procedure since it denervates the parietal cell mass but leaves the antral mill (antropyloric segment) innervated and therefore obviates the need for a drainage procedure. Herein lies the distinct advantage of this operation as the avoidance of drainage leads to virtual abolition of the alimentary side effects, although diarrhoea can still occur, albeit extremely rarely. The drawback of HSV is the higher incidence of recurrent ulceration documented by some long-term reports (20–30%). These reports have, however, included cases for which HSV is nowadays considered unsuitable. Thus HSV should not be performed in patients with pyloric channel ulcers (including prepyloric), in patients with stenosis (previously these were managed with HSV and pyloric dilatation) and in patients with bleeding ulcers. HSV is easily performed by the laparoscopic approach.

Alternatives to HSV include the Taylor procedure (posterior truncal vagotomy and anterior seromyotomy) and the Hill procedure (posterior truncal vagotomy and anterior highly selective vagotomy). Several reports and one clinical trial have demonstrated that the Taylor procedure gives equivalent results to the classical HSV in the short term but there are no long-term reported series. There have been no large series or clinical trials with the Hill procedure, although on *a priori* grounds it should achieve equivalent results. Both the Taylor and the Hill operations can easily be performed laparoscopically and there are several such reports on the laparoscopic Taylor operation.

Drainage procedure

Truncal vagotomy and bilateral selective vagotomy necessitate a drainage procedure. This can be achieved by either a



	Subtotal gastrectomy	Vagotomy and drainage	Highly selective vagotomy
Mortality	2%	1%	0.5%
Recurrence rate	2–3%	5–8%	5–30%
Side effects	+++	++	±

Figure 23.15 The relative advantages and disadvantages of the various procedures used for chronic duodenal ulceration. The mortality and ulcer recurrence data are average values only. Choice of the 'best operation' for the individual patient involves a careful balance of these three factors.

gastrojejunostomy (anterior or posterior) or a pyloroplasty (Heineke–Mikulicz or Finney). Most surgeons would prefer the posterior short loop to the anterior (longer) gastrojejunostomy although there are no comparative trials. With Heineke–Mikulicz pyloroplasty, the pylorus is incised longitudinally and then sutured transversely, whereas the Finney procedure consists of an anastomosis between the duodenum and the antrum through an inverted U-shaped incision. There is not much to choose between these two in the absence of duodenal scarring, in which case the Finney pyloroplasty is preferable.

Controlled clinical studies comparing gastrojejunostomy vs pyloroplasty have not indicated any obvious superiority between these two forms of drainage, but there are some important practical considerations. Bile vomiting is slightly commoner after gastrojejunostomy but this is offset by a slightly higher incidence of dumping after pyloroplasty. Gastrojejunostomy is a safer option when the ulcer area is the seat of an inflammatory mass or when there is significant fibrous scarring. If a patient develops bile vomiting or dumping after vagotomy and drainage, it is easy to close the gastrojejunostomy. In contrast, reconstruction of the pylorus is more technically demanding and the results of remedial surgery less certain.

Operations for gastric ulcer

The objectives of gastric ulcer surgery are the removal of the ulcer, thus dealing with the problem of possible malignancy, and the prevention of further recurrence. Ideally, therefore, the ulcer should be removed as part of the gastrectomy specimen, but as an alternative local excision of the ulcer is combined with vagotomy. The standard treatment is the Billroth I gastrectomy. This resection includes the ulcer, the lesser curve and the antrum. Restoration of continuity by gastroduodenal anastomosis is seldom a problem as the duodenum is normal. Both the mortality and the ulcer recurrence rate following this operation are low (2%) and it is currently the gold standard procedure for the surgical treatment of gastric ulcer disease. Postoperative alimentary side effects can follow this operation but their frequency and severity are less than those encountered after Pólya gastrectomy or truncal vagotomy with drainage. Likewise, long-term nutritional problems are not usually marked or frequent after Billroth I gastrectomy.

The Pólya gastrectomy (also sometimes referred to as Billroth II) is less popular nowadays because of the associated higher incidence of postgastric surgery syndromes. The resection is more extensive and includes the first part of the duodenum with closure of the duodenal stump and an end-to-side gastrojejunostomy. This operation was popular in the past particularly for the emergency treatment of a bleeding pyloric channel and duodenal ulcers.

Nowadays, the main alternative to Billroth I gastrectomy in the treatment of gastric ulcer is vagotomy and drainage. This carries a lower operative mortality (1%). If this operation is preferred, it should be combined with local excision of the ulcer whenever possible. A full thickness biopsy is mandatory if ulcer excision is not possible because of its location. The recurrence rate after truncal vagotomy and drainage with excision of the gastric ulcer is high and averages 10–20% in the long term. Furthermore, the symptomatic results, in terms of alimentary problems, are no better than those after Billroth I gastrectomy, though better than after Pólya gastrectomy. There is therefore little justification in using this operation as the routine surgical treatment for uncomplicated type I gastric ulcer. It is, however, indicated in the poor-risk elderly patient. The reported experience of HSV with ulcer excision is limited but the results appear to approximate those achieved by truncal vagotomy and drainage, although one report documented an unacceptably high ulcer recurrence rate.

A special problem is posed by the high lesser curve ulcer situated at the cardio-oesophageal junction. The options here include a resection distal to the ulcer, leaving it *in situ* (Kelling–Madlener procedure) or, preferably, to excise all the lesser curve using the Pauchet type of gastrectomy where the new lesser curve is reconstituted without the use of clamps. Unfortunately, very high lesser curve ulcers are often encountered in elderly patients where the questions of malignancy and operative mortality are especially pertinent. Proximal or total gastrectomy is inadvisable in these patients and, whenever possible, medical therapy should be persisted with. An HSV with local excision of the ulcer merits consideration in elderly patients who prove refractory to medical treatment.

Some 30% of patients with lesser curve ulceration are said to have evidence of past or present duodenal ulceration. Such ulcers have been designated as type III. It is usually assumed that the duodenal ulcer precedes the gastric, but the precise relationship is not always clear. In theory, both delayed gastric emptying from pyloric stenosis or bile reflux through an incompetent diseased duodenal bulb could be responsible for the gastric ulcer.

Gastric ulcers coexisting with duodenal ulcer disease tend to be located fairly distally in the stomach, usually at or beyond the incisura. Levels of gastric acid secretion are normal or high but below levels found in patients with duodenal ulcer disease alone. Surgical treatment of combined ulcer disease is by the operation normally favoured by the surgeon for duodenal ulcer. The risk of gastric ulcer being malignant when it is associated with duodenal ulcer is small but, nonetheless, present. In one large series, gastric carcinoma was present in 8% of patients with gastric ulcer alone and 1.2% of patients with coexisting duodenal ulceration. All forms of gastric ulcer must be biopsied.

Laparoscopic elective ulcer surgery

All elective operations for peptic ulcer disease can and are being carried out laparoscopically with good results in terms of immediate outcome although there are no published long-term studies and thus data on completeness of vagotomy and ulcer recurrence rates compared with the conventional open approach are not available. What is established from the retrospective published series is that the immediate clinical outcome is the same in terms of ulcer healing and that the laparoscopic approach is attended by a low incidence of wound-related complications, reduced postoperative pain and period of ileus, resulting in a shorter hospital stay and accelerated recovery to full activity or work. Thus the laparoscopic approach seems ideal for the treatment of failures of medical therapy for both duodenal ulcer (usually HSV) and gastric ulcer (Billroth I gastrectomy). The laparoscopic approach is equally appropriate for the emergency treatment of complicated peptic ulcer disease, especially perforation. Indeed judging by the increasing number of published series and randomized clinical trials, laparoscopic treatment of perforated peptic ulcer disease is likely to replace the open approach as, aside from the general benefits of the minimal access approach, it is accompanied by a lower incidence of surgical site infections and wound-related complications.

Failures of gastric surgery

Unsatisfactory results follow ulcer surgery in 10–20% of patients. The reported incidence varies considerably and no doubt depends to some extent on the thoroughness and length of follow-up. Poor results may be due to recurrent ulceration, alimentary side effects leading to various postgastric surgery syndromes or adverse nutritional consequences. Thus the failures of gastrectomy are largely nutritional or due to postcibal or alimentary sequelae, and rarely caused by recurrent ulceration; with the opposite being the case after HSV. The poor results of truncal vagotomy with drainage may be due either to recurrent ulceration or to functional postcibal side effects.

Recurrent ulcer

Recurrence of type I gastric ulcer after an appropriate gastrectomy is rare, so the problem concerns mainly duodenal and pyloric channel ulcers (type II). Recurrent ulceration may appear in the stomach, the duodenum, or at the site of the gastrointestinal stoma (anastomotic ulcer). The overall rate depends on the procedure performed, the ulcer type and personal habits of the patient, e.g. smoking, alcohol, etc. Thus the highest recurrent ulcer rates are reported after HSV and the lowest after truncal vagotomy and antrectomy. Prepyloric ulcers exhibit a higher recurrence rate than duodenal ulcers. A Swedish study has demonstrated that continued smoking is an important factor and resulted in a recurrence rate of 24% as opposed to 7% in non-smokers after HSV. Several studies over the years have confirmed that the two most important factors are incomplete vagotomy (denervation of the parietal cell mass) and inadequate drainage (gastric hold-up) and these probably account for 80% of recurrences. It should be stressed, however, that these data on recurrent ulceration after surgical treatment predate the *H. pylori* era, and that it seems highly likely that many of these patients with recurrent ulcers harboured this infection. Thus the true recurrence after surgical treatment in *H. pylori*-negative patients is not known.

Table 23.3 Laboratory findings associated with recurrent ulcer after surgery

Test	Result
Helicobacter pylori breath test	Positive
Basal acid output	>4 mmol/h
Postoperative % reduction of basal acid output	<70%
Maximal acid output	>25 mmol/h
Insulin test	Positive
Sham feeding test	Positive
Basal serum group I pepsinogen	>100 ng/mL
Betazole-stimulated group I pepsinogen	>100% of basal

The laboratory tests needed in the assessment of patients with recurrent ulcer are shown in Table 23.3. In patients with recurrent ulcer after vagotomy with drainage or antrectomy, the basal serum pepsinogen I (PG I) levels are significantly higher than in patients without recurrence.

Furthermore, following injection of betazole, patients with recurrent ulcer exhibit a rise in the serum PG I level whereas those without (or with a complete vagotomy) demonstrate a paradoxical decrease in the serum level of PG I. Rare causes of recurrent ulceration include inadequate gastric resection, retained antrum, Zollinger–Ellison syndrome and hypercalcaemia.

Inadequate vagotomy

Failure to achieve complete denervation of the parietal cell mass is the commonest cause of recurrent ulceration. The incidence of incomplete vagotomy may be as high as 20% after truncal vagotomy, as judged by the results of the early postoperative insulin test. It is true that some of these positive results are only technically positive and may be the result of very small residual fibres, but some are due to a missed major vagal trunk, in which case recurrence is inevitable. Preoperative tests have been described which are reputed to improve visualization of the nerve fibres (leucomethylene blue) or permit a check on the completeness of the vagotomy by observing a failure of peristalsis or absence of acid secretion after electrical stimulation or after pentagastrin or histamine infusion (as shown by an intragastric dye such as Congo red or pH probe). In practice, these tests have not become established in routine practice. The most important prerequisites for successful (complete) vagotomy are an appreciation of the variable anatomy of the vagi, patience and meticulous technique.

Truncal vagotomy can be carried out at two levels: the hiatus or above (as advocated by Dragstedt), or at the cardia. At either level, it is possible to miss vagal fibres (Figure 23.16), although there is less likelihood to overlook a significant trunk if the vagotomy is carried now down on the oesophagus. The solution to the problem is to divide the anterior and posterior trunks at the hiatus initially. This will enable the oesophagus to be pulled down and the region of the cardia then becomes more accessible. The lower

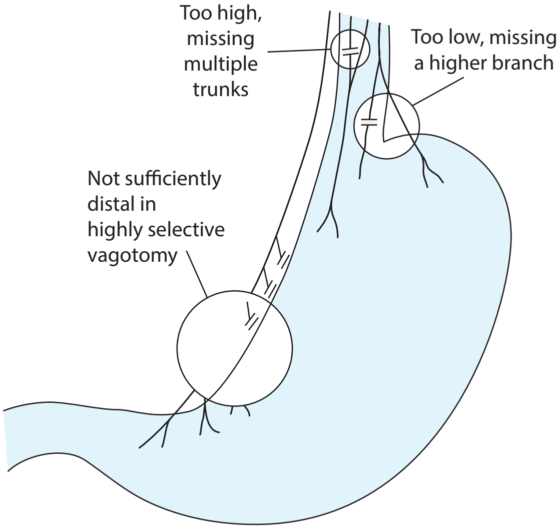

Too high, missing multiple trunks
Too low, missing a higher branch
Not sufficiently distal in highly selective vagotomy

Figure 23.16 Causes of incomplete vagotomy.

abdominal oesophagus is then cleared methodically all the way round of all nerve fibres. With this length of abdominal oesophagus cleared, it is quite easy to inspect the region of the fundus and thus ensure against missing nerve fibres (of Grassi) to this region.

If complete denervation of the parietal cell mass is to be achieved, three aspects of the technique need special attention. Clearance of at least 5 cm of the lower abdominal oesophagus is mandatory. In addition, the denervation of the lesser curve must be extended distally enough – at least 2.5 cm from the oesophagogastric junction. In HSV, this means leaving only one recognizable anterior antral branch of the nerve of Latarget. Finally, the innervation of the fundus by the nerves of Grassi must be disrupted.

Antral stasis

This arises when there is duodenal fibrosis. Previously these cases were treated by pyloric dilatation during the HSV. Undoubtedly this has been one of the factors involved in ulcer recurrence after this procedure. Nowadays, it is recognized that these patients require a pyloroplasty. The nature of the vagotomy (HSV or truncal) then becomes a matter of individual choice. Another unsatisfactory practice is the performance of bilateral truncal vagotomy with pyloric stretch instead of drainage. The ulcer recurrence (usually pyloric channel) in these patients with time is unacceptably high.

Rare causes of ulcer recurrence

These include inadequate gastric resection, retained gastric antrum, the Zollinger–Ellison syndrome and hypercalcaemia (Figure 23.17).

Inadequate gastric resection

If gastrectomy is the procedure used for reducing acid secretion, a 75% resection is needed. A related consideration is the method of restoring continuity after resection. In terms of ulcer recurrence, it is generally believed that a gastrojejunal anastomosis (Billroth II, Pólya) is better than gastroduodenostomy, although it is difficult to substantiate this claim from the published literature. The difference in ulcer recurrence rates may be related to the extent of gastric resection since more stomach is likely to be retained in Billroth I operations to achieve a gastroduodenal anastomosis without tension.

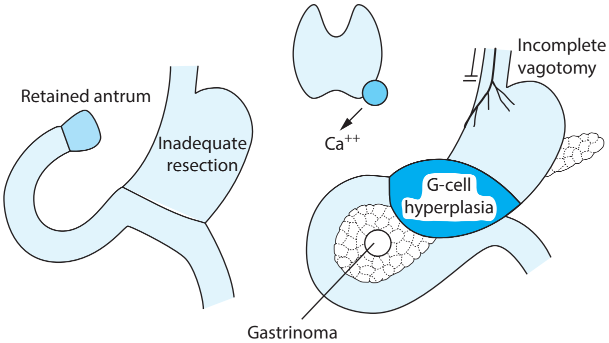


Figure 23.17 Rare causes of recurrent ulceration after surgical treatment of peptic ulcers.

Retained gastric antrum

If antral tissue (after gastric resection) is left in continuity with the duodenal stump, the antral G-cells now in a permanent alkaline environment will secrete gastrin continuously. Under this constant stimulus, the residual parietal cells will produce sufficient acid for stomal ulceration to develop. Whenever hypergastrinaemia is found after partial gastrectomy, the retained antrum syndrome must be considered. Differentiation of this condition from autonomous tumour production of gastrin is usually possible from the results of the secretin and calcium tests, but direct confirmation of retained antral tissue may be more difficult. Radioisotope scanning with pertechnetate may help. Treatment entails excision of the retained antral tissue.

Zollinger–Ellison syndrome (gastrinomas)

This was first described in 1955 in a classical surgical paper by Zollinger and Ellison before the isolation of gastrin by Gregory and Tracey in a 36 year old female patient with severe recurrent ulcer disease and multiple pancreatic tumours who was cured of her ulcer diathesis by a pancreaticoduodenectomy. It is now established that Zollinger–Ellison syndrome is a clinical syndrome with severe peptic ulcer disease often associated with erosive oesophagitis and sometimes diarrhoea caused by gastric acid hypersecretion secondary to a neuroendocrine tumour (NET) that secretes excessive amounts of gastrin. Zollinger–Ellison syndrome is now recognized to consist of two types – sporadic and hereditary–and is the presenting feature of *multiple endocrine neoplasia type 1* (MEN-1, Werner syndrome), which is described in Chapter 27. The definition of Zollinger–Ellison syndrome outlined above is important as some NETs although expressing gastrin are not functionally active. The term gastrinoma is nowadays reserved to a tumour secreting gastrin autonomously and thus causing the Zollinger–Ellison

BOX 23.3 Relative features of sporadic and hereditary Zollinger–Ellison syndrome

- Sporadic cases account for up to 75% of cases
- Clinical features of the two types are identical
- Zollinger–Ellison syndrome is the presenting feature in 40–60% of MEN-1
- Over 90% of tumours in Zollinger–Ellison syndrome with MEN-1 are located in all parts of the duodenum with highest incidence in first part, and often multiple and small (<10 cm) and thus difficult to detect
- In sporadic Zollinger–Ellison syndrome , the gastrinomas may be located in either the duodenum or pancreas - the latter are usually solitary and bigger, with the majority located in the head
- The prognosis of sporadic and MEN-1-associated duodenal gastrinomas is better than that of pancreatic gastrinomas, as the duodenal tumours progress slowly to liver metastasis

syndrome. The relative features of the sporadic and hereditary Zollinger–Ellison syndrome have been reported in detail and are summarized in Box 23.3.

In one of the largest registries of sporadic NETs of the pancreas (535) collected from the archives of the departments of pathology in Zurich, Switzerland, and Kiel, Germany, 24 patients (4.5%) suffered from sporadic pancreatic gastrinomas and Zollinger–Ellison syndrome. An additional 19 patients suffered from MEN-1 and Zollinger–Ellison syndrome. The latter group of patients showed exclusively duodenal gastrinomas, with no pancreatic tumours. The incidence of sporadic duodenal gastrin-producing tumours appears to be increasing, but this is likely to be due to improved diagnosis. In sharp contrast, pancreatic MEN-1-associated gastrinomas are very rare if one excludes the large subset of tumours with immunohistochemical expression of gastrin but without evidence of Zollinger–Ellison syndrome. These are now designated as *functionally inactive NETs expressing gastrin*, rather than gastrinomas. Sporadic gastrinomas account for 70% of gastrinomas and are malignant but usually of low grade in 40–85%. They occur in a region described as the *gastrinoma triangle* – the confluence of the cystic with the common bile duct superiorly, the second and third portions of the duodenum inferiorly, and the neck and body of the pancreas medially. Duodenal sporadic gastrinomas arise in the submucosa and may be multiple whereas sporadic pancreatic gastrinomas tend to be solitary. Although metastasis to lymph nodes occurs early, hepatic metastases occur late and only in 10% of patients. Primary gastrinomas may very rarely occur in other sites: body of the stomach, jejunum, peripancreatic lymph nodes, splenic hilum, root of the mesentery, omentum, liver, gallbladder, common bile duct, and the ovary. Primary lymph node gastrinomas (disputed by some and considered as deposits from a small undetectable primary duodenal gastrinoma) are said to be due to entrapment of neuroendocrine cells during development. Whatever the truth between these opposing views, apparently nodal disease without an obvious other tumour is encountered in 5–10% of sporadic Zollinger–Ellison syndrome.

Signs and symptoms of the sporadic and hereditary Zollinger–Ellison syndrome do not differ. The hypergastrinaemia by inducing continuous gastric acid hypersecretion leads to mucosal ulcerations in the oesophagus, stomach, duodenum and even jejunum. Some 20% of patients exhibit secretory diarrhoea alone or in combination with upper gastrointestinal ulceration. Abdominal pain is a frequent symptom (>75% of the patients). The important differences from the usual peptic ulcer disease include multiple ulcers, ulcers in atypical locations, persistence and/or difficult control of symptoms despite adequate antacid therapy and diarrhoea. All these unusual features should suggest the possibility of Zollinger–Ellison syndrome and the need for appropriate testing. Despite this, the mean interval from first symptoms to definite diagnosis is usually long, up to 4 or more years. This may be attributed to late presentation in some patients either with a mild form of peptic ulcer disease or with secretory diarrhoea.

The diagnosis is confirmed by serum gastrin assay, and a gastrin concentration >1000 pg/mL is diagnostic. However, the levels are lower in some 60% of patients and the issue is compounded by the high prevalence of patients on chronic PPI medication causing hypergastrinaemia. In these borderline cases, the secretin test is indicated. It is considered positive (diagnostic) if it increases the serum gastrin by more than 200 pg/mL. It is recommended that the PPI should be stopped for 1 week before a secretin test is performed, but as this may put the patient at risk from complications close observation is essential during this period.

Once the diagnosis is established the patient is immediately put on a high-dose PPI, which is nearly always effective. One large review of 80 patients confirmed that high-dose omeprazole (varying from 60 to 120 mg) daily provided satisfactory control in 90% of patients. The important practical issue is that the dose of PPI has to be individualized as the response to PPIs varies from patient to patient. Localization of the gastrinoma and imaging tests to stage the disease (MDCT, MRI, PET-CT) are used mainly to detect nodal involvement and hepatic metastases. Somatostatin receptor scintigraphy (SRS) with ^{111}In-labelled octreotide is used for the detection of gastrinomas as most of these tumours overexpress the somatostatin receptor subtype 2. Published reports show a high yield of SRS in the localization and staging of gastrinomas (80%), and SRS is regarded as being superior to other non-invasive imaging tests and contributes to patient management. The problem relates to its sensitivity, as this depends on the size and exact localization of the tumours. Thus small pancreatic lesions and lesions located in the duodenum may be missed.

The real controversy lies in the definitive treatment of Zollinger–Ellison syndrome and, in particular, the role of surgical treatment of the disease. There is little doubt that long-term PPI therapy is highly effective and safe especially as the disease, even when malignant, is compatible with long-term survival, and indeed the most important prognostic factor in the individual patient is the biological aggressiveness of the NET. Certainly, patients with gastrinoma associated with type 1 (MEN-1) are seldom if ever cured by non-Whipple operations, i.e. by removal (enucleation) of duodenal and pancreatic gastrinomas. Most surgical experts consider that surgery should be used in these patients to deal with malignant pancreatic or duodenal tumours, the malignant potential being indicated by the size of the tumour. According to this policy, tumours greater than 2 cm in size warrant excision. Cure rates as high as 50% have been reported after local excision of malignant gastrinomas. In this respect, wide duodenotomy is essential for improved tumour detection cure rates. In experienced hands, and in the absence of hepatic deposits, Whipple pancreaticoduodenectomy results in the highest probability of cure in both sporadic and hereditary type MEN-1 gastrinoma patients as it removes the entire gastrinoma triangle. However, this has to be balanced against the excellent long-term survival of these patients with lesser operations, especially in view of the increased operative mortality and long-term morbidity of the Whipple procedure. There is no indication for total gastrectomy nowadays as inoperable disease is controlled by various forms of drug therapy in addition to PPIs.

In one-third of the patients the tumour has already metastasized to the liver by the time of diagnosis. The palliative treatment options in these patients include chemotherapy, peptide radioreceptor therapy (PRRT) and biotherapy with somatostatin analogues and is indicated if there is evidence of disease progression. PRRT is based on the established fact that well-differentiated NETs overexpress somatostatin receptors which are targeted by this form of radionuclide therapy. There is emerging evidence that PRRT may be especially effective in gastrinomas and cases when sufficient downstaging by PRRT enabled subsequent resection are documented. Usually, however, this treatment is considered palliative and several reports have documented partial response or disease stabilization. Biotherapy with somatostatin analogues is also used in progressive symptomatic inoperable well-differentiated disease (G2), usually before recourse to chemotherapy with streptozotocin plus doxorubicin/5-fluorouracil (5FU). In poorly differentiated gastrinomas (G3) chemotherapy is usually with etoposide and cisplatin. Future therapeutic strategies are predicted to include angiogenesis inhibitors and kinase inhibitors targeting tumour-specific signalling cascades.

Antral G-cell hyperplasia (pseudo-Zollinger–Ellison syndrome or Zollinger–Ellison syndrome type 1)

This condition is referred to as Zollinger–Ellison syndrome type 1 to distinguish it from the classical Zollinger–Ellison syndrome associated with a gastrinoma and referred to as Zollinger–Ellison syndrome type 2 in this context. It is a very rare condition where the hypergastrinaemia results from an increased number of antral gastrin-secreting cells or from an exaggerated response to feeding (postprandial hypergastrinaemia without hyperplasia). The syndrome is therefore better referred to as antral G-cell hyperactivity or hyperfunction. The recommended treatment is antrectomy. Some doubt the existence of this syndrome and cite the absence of reported cases in recent years as evidence. The controversy is compounded further by the undoubted existence of rare cases of recurrent and intractable peptic ulceration in association with a pancreatic endocrine tumour that does not secrete gastrin although the nature of the factor has yet to be identified.

Hypercalcaemia

Primary hyperparathyroidism can, on rare occasions, cause peptic ulceration in the absence of a gastrinoma. These patients do not usually have hypergastrinaemia. In addition, hyperparathyroidism is, of course, part of the MEN-1 syndrome associated with Zollinger–Ellison syndrome, in which case the serum gastrin will be abnormally high. In practice, if hypercalcaemia is found in a patient with peptic ulcer, recurrent or otherwise, both parathormone and gastrin levels should be assayed. If primary hyperparathyroidism only is established by this testing and corrected surgically, the ulcer should heal. There is some evidence that this is more likely to occur if the patient has a parathyroid adenoma than hyperplasia. If MEN-1 Zollinger–Ellison syndrome is established treatment is that of gastrinoma takes precedence.

Management of patients with recurrent dyspepsia

The persistence of indigestion, or its return after a period of freedom from symptoms, suggests the possibility of recurrent ulceration. Other possibilities should not, however, be overlooked. These include unrelated pathology such as gastro-oesophageal reflux disease, gallstones and pancreatitis. On other occasions, the dyspepsia is caused by complications of ulcer surgery, e.g. enterogastric reflux or gastric stasis. It is not always easy to diagnose recurrent ulceration on clinical grounds alone. Sometimes, the character of the pain is similar to that experienced by the patient before operation, but on other occasions it is different in type, site or both. If vomiting is a feature, the presence of bile or food may suggest gastritis or stasis as the underlying cause of the patient's symptoms. Recurrent ulcers may present with bleeding or, less commonly, perforation with no accompanying dyspepsia. The rapid return of symptoms, haemorrhage or perforation, particularly in the younger patient should suggest the possibility of Zollinger–Ellison syndrome. Coexisting diarrhoea may be indicative of this condition or a gastrojejunocolic fistula, although simple postvagotomy diarrhoea is much more common than either of these two conditions. Gastrojejunocolic fistula is an extremely rare complication of benign ulcer disease nowadays and the vast majority of cases are neoplastic in origin.

Following gastric surgery, radiological contrast studies of the stomach and duodenum though occasionally useful are often not, and are difficult to interpret. Scarring of the duodenum will persist even if the ulcer has healed and the presence of a pyloroplasty makes radiological identification of a recurrent ulcer in this region virtually impossible. Endoscopy with assessment of the *H. pylori* status is the best method for the investigation of patients with recurrent dyspepsia and, especially, for the detection of ulcer recurrence. It may also provide additional information relating to adverse consequences of gastric surgery, particularly enterogastric bile reflux and gastritis. The diagnosis of recurrent ulceration may (rarely nowadays) be followed by investigation to assess the postoperative gastric acid secretion. These tests will provide information on the completeness or otherwise of the vagotomy and the persistence of acid hypersecretion which, if excessive, should raise the possibility of Zollinger–Ellison syndrome.

Estimation of serum gastrin is wise in all patients with recurrent ulcer. If hypergastrinaemia is detected, further investigations are required to exclude gastrinoma.

Not all recurrent ulcers require surgical treatment. Indeed the prevalent view is that the vast majority should be treated with a full 2 month course of PPI with eradication in those patients who test positive for *H. pylori* by the ^{13}C-urea breath test. The specific indications for surgical intervention include antral stasis associated with pyloric channel ulcers, incomplete vagotomy and retained antrum (following gastrectomy). Patients with incomplete vagotomy who are *H. pylori* negative are best treated with revagotomy and antrectomy.

BOX 23.4 Sequelae of gastric surgery

Recurrence of the disease

Nutritional and functional sequelae
- Weight loss, anaemia: iron deficiency and B_{12} deficiency
- Milk intolerance
- Bone disease
- Dumping syndromes
- Reactive hypoglycaemia
- Bile vomiting
- Diarrhoea
- Small stomach syndrome

Mechanical complications
- Afferent/efferent loop obstruction
- Jejunogastric intussusceptions
- Gastro-oesophageal reflux

Others
- Cholelithiasis
- Bezoar formation
- Gastric carcinoma

Sequelae of gastric surgery

Minor postprandial complaints are commonly experienced by patients after gastric operations. These usually improve with time and dietary adjustments. In a cohort of patients, however, variously estimated at 5–20%, the symptoms are severe, persistent and cause considerable disability, often drastically reducing the quality of life, and malnutrition. The problem may affect the quality of life so adversely that the patient commits suicide. This regrettable outcome was encountered in two patients in a series of 321 patients with postgastric surgery problems referred to the author for remedial treatment/surgery over a period of 30 years. It seems likely from recent reports that similar problems are being encountered in patients after gastric bypass, pancreatic biliary bypass and duodenal switch operations for morbid obesity.

The various postgastric surgery syndromes arise on a background of altered anatomy and physiology of the upper gastrointestinal tract, although the exact mechanisms responsible for some of the severe symptoms remain unclear. A useful classification of the sequelae of gastric surgery is shown in Box 23.4.

In the author's personal series, no particular personality type is at increased risk, but disabling symptoms after gastric surgery are more commonly encountered in the following:

- female sex
- operations for peptic ulceration in the young (below 30 years)
- extensive gastrectomy with duodenal diversion (Pólya).

Severe and persistent symptoms are rarely encountered after HSV if one excludes symptoms caused by recurrence of the ulcer. However, they occur with similar frequency to that reported after gastrectomy in patients who undergo truncal vagotomy with drainage or truncal vagotomy and antrectomy. The type of drainage procedure (pyloroplasty or gastrojejunostomy) does not affect the incidence of postprandial symptoms and other sequelae.

Nutritional consequences of gastric surgery

These consist of weight loss, anaemia and bone disease.

Weight loss

Loss of weight or failure to gain weight is very common after gastric surgery and tends to be more marked after extensive gastrectomy, particularly of the Pólya type. Often there is early satiety due to loss of adaptive relaxation and disturbances of ghrelin release. In practice, significant weight loss is usually encountered in patients who obtain a bad functional result and who experience severe postcibal symptoms such that they are afraid to eat. The resulting diminished calorie and protein intake is the major factor, although malabsorption of fat and nitrogen and decreased small bowel transit time may be operative, at least in some patients. Although mild steatorrhoea is common, severe fat malabsorption is rare unless there is a coexisting subclinical small bowel disease (e.g. gluten enteropathy) or gross bacterial overgrowth.

Anaemia

Iron-deficiency anaemia

Microcytic hypochromic anaemia is very common after vagotomy and drainage and gastric resections, especially in females. The incidence of this complication increases with time and approximates to 60% and 80% at 10–20 years in males and females respectively. The exact pathogenesis of the iron-deficiency anaemia is unclear but is probably multifactorial. The mechanisms thought to be important include:

- shift to trivalent ferric iron at high pH followed by polymerization
- loss of a gastric juice factor which normally facilitates the absorption of iron
- diminished splitting of iron–protein complexes by the reduced peptic activity of the gastric juice
- enhanced binding of dietary iron to specific proteins (e.g. gastroferrin).

In view of the high incidence of iron-deficiency anaemia after gastric surgery, prophylactic treatment with oral iron (300 mg q.d.s.) is nowadays recommended in all patients after gastrectomy and truncal vagotomy with drainage or antrectomy. This amount of daily iron supplementation allows sufficient absorption to restore serum iron levels to normal.

Macrocytic anaemia

This is the result of vitamin B_{12} deficiency. Malabsorption of this vitamin is invariable after total gastrectomy due to the loss of IF. However, megaloblastic anaemia takes several years to develop due to the large body stores of vitamin B_{12}.

These patients have an abnormal Schilling test and require 3 monthly injections of cyanocobalamin indefinitely. Subclinical deficiency of this vitamin is also encountered in some patients after partial gastrectomy and vagotomy with drainage or antrectomy, although frank megaloblastic anaemia is rare in these groups. The main factor responsible for the impaired absorption of dietary vitamin B_{12} in patients after partial gastrectomy and truncal vagotomy is the lack of acid environment which normally facilitates the release of vitamin B_{12} bound to ingested food. The reduced secretion of IF reported in some patients is considered to be less important in this group of patients in whom the Schilling test is normal. Treatment is with oral crystalline vitamin B_{12}, which is administered between meals. Malabsorption of vitamin B_{12} may also be the consequence of bacterial overgrowth and steatorrhoea. Folate deficiency is rare and is only encountered in patients after extensive or total gastrectomy. It results from an inadequate dietary intake,

Bone disease

This complication develops several years after gastric resection with duodenal exclusion (Pólya) as the duodenum is the major site of calcium absorption. The majority of patients are females who classically develop osteomalacia 10–20 years after the gastrectomy. In fact, in the vast majority of cases features of both osteomalacia (demineralization of bone) and osteoporosis (loss of bone substance) are present. The biochemical features (raised alkaline phosphatase and serum calcium) usually predate the clinical symptoms by several years. The clinical features of postgastrectomy bone disease include generalized bone pains, weakness due to an associated myopathy and the development of stress fractures most commonly in the vertebral column. Treatment should be prophylactic with oral calcium and vitamin supplements. The established condition may be treated with bisphosphonates.

The dumping syndrome

Considerable confusion has been generated by the inclusion of patients with reactive hypoglycaemia in the group and referred to as 'late dumping' to differentiate this condition from patients with vasomotor symptoms which occur soon after eating and in this erroneous classification are designated as 'early dumpers'. There is now general agreement that patients with symptoms due to reactive hypoglycaemia which occur 2–3 hours after a meal should not be included in the dumping syndrome.

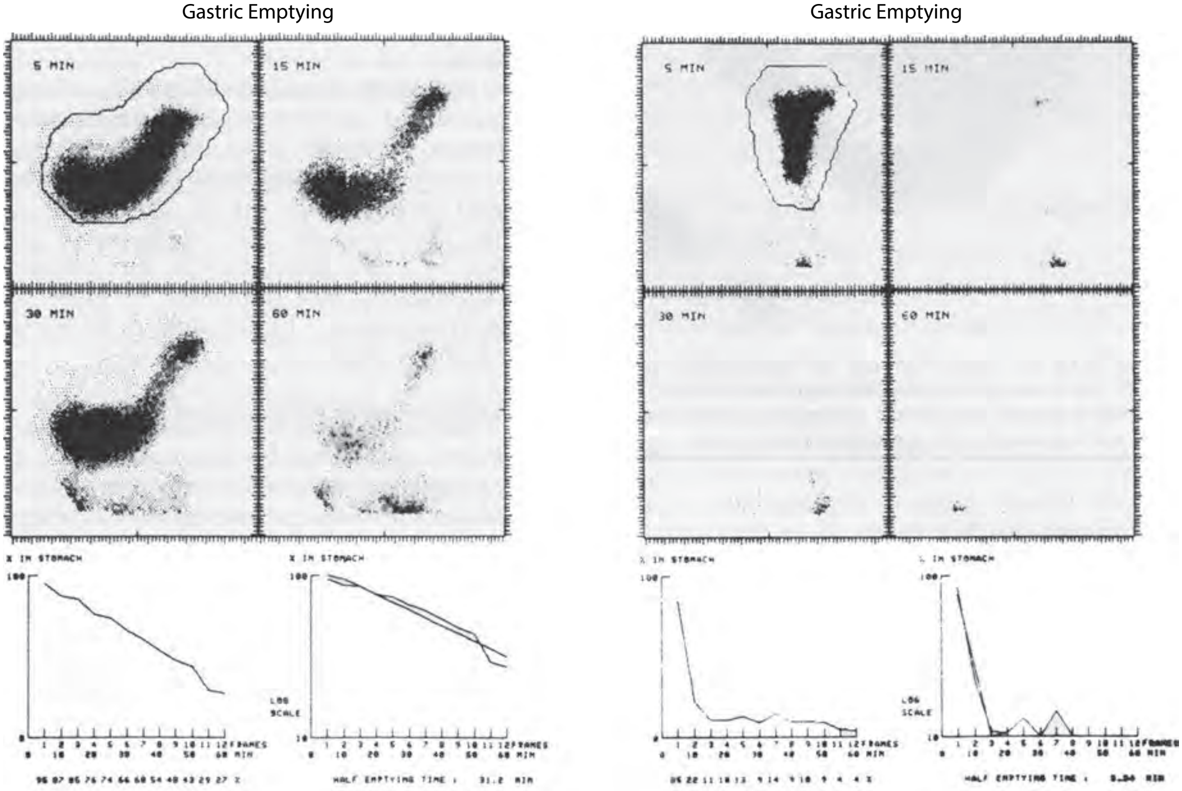
Although the term 'dumping' was introduced by Mix in 1922, the first description of the symptoms of vasomotor dumping syndrome had been reported previously by Hertz in 1913. The syndrome, which is one of the commonest sequelae of gastric surgery, consists of postprandial vasomotor (systemic) and gastrointestinal symptoms (Table 23.4). The dumping syndrome is associated with rapid gastric emptying (Figure 23.18), although some have postulated that the enterogastric reflux of bile is responsible for some of the symptoms. The fact is that both enterogastric reflux of bile and vasomotor dumping may occur in the same patient, and other than sharing the same aetiology (gastric surgery) there is no evidence that one syndrome causes the other.

The vasomotor symptoms and signs (palpitations, vasodilatation, hypotension and fainting/having to lie down, etc.) occur within minutes of eating and are principally caused by hypovolaemia, which is accompanied by diminished cardiac output and reduced peripheral resistance. The attacks are typically precipitated by high-carbohydrate meals. The hypovolaemia is secondary to a massive outpouring of fluid from the vascular compartment into the small bowel lumen as a consequence of the hyperosmolar nature of the intestinal contents resulting from the precipitous gastric emptying. Several vasoactive peptides have been held responsible as mediators of the vascular and gastrointestinal manifestations of the dumping syndrome. These include kinins, substance P, enteroglucagon, gastric inhibitory polypeptide and neurotensin. The gastrointestinal symptoms, which include diarrhoea, occur later during the course of a dumping attack and may be absent.

Patients with mild to moderate dumping symptoms are managed satisfactorily with dietary manipulations. These patients are advised to eat small, dry meals (no fluids before or during meals) rich in protein and fat but low in carbohydrates. Additives that slow gastric emptying, such as methoxy-pectin or bran,

Table 23.4 Manifestations of the dumping syndrome

Vasomotor (systemic)	Gastrointestinal
Weakness	Fullness and early satiety
Tiredness	Epigastric discomfort/heaviness
Dizziness	Nausea
Headache	Vomiting
Fainting/wanting to lie down	Distension
Feeling of warmth	Excessive borborygmi/distension
Dyspnoea	Diarrhoea
Sweating	



(a) (b)

Figure 23.18 Gastric emptying of an isotope-labelled meal: (a) normal single exponential emptying; (b) rapid initial gastric emptying in a patient with severe dumping symptoms.

are beneficial. However, remedial gastric surgery is required for patients with severe and persistent dumping (see below).

Reactive hypoglycaemia

This complication is relatively uncommon and has a reported incidence of 1–6% after gastric surgery. Reactive hypoglycaemia often exists with other symptoms, including vasomotor dumping and diarrhoea. The symptoms, which occur 2–3 hours after a meal, are due to hypoglycaemia and include sweating, tremor, difficulty in concentration and, rarely, fainting. The diagnosis is best confirmed by an extended oral glucose tolerance test, which demonstrates an initial hyperglycaemia. This triggers an exaggerated insulin release with elevated plasma insulin and enteroglucagon that precede the hypoglycaemia.

Reactive hypoglycaemia usually responds to dietary measures, including small low-carbohydrate, high-protein meals and very rarely requires remedial surgical treatment.

Bile vomiting

Vomiting of bile or bile-stained fluid before or after meals may be a manifestation of the following disorders:

- recurrent ulceration
- enterogastric reflux
- intermittent obstruction of the afferent or efferent loop of a gastrojejunostomy
- cardio-oesophageal incompetence.

Enterogastric reflux/reflux gastritis

Reflux of upper intestinal secretions (bile/pancreatic juice/succus entericus) into the stomach causes a reflux erosive gastritis and bile vomiting. The symptoms include epigastric pain, nausea and vomiting in the early postprandial period. The pain is usually of a burning nature, is aggravated by food and is not relieved by antacids. The attack usually culminates in the vomiting of bile-stained fluid 1–2 hours after a meal. Less commonly, the vomiting occurs in the early morning and is preceded by nocturnal burning pain. The erosive gastritis leads to chronic blood loss with the development of an iron-deficiency anaemia and, occasionally, to overt acute gastric haemorrhage.

The diagnosis is established by upper gastrointestinal endoscopy, which shows a diffuse gastritis with oedematous friable mucosa and superficial erosions, in addition to pooling of bile-stained fluid. Quantification of the enterogastric reflux is obtained by the modified EHIDA test. In this investigation, EHIDA is injected intravenously and is followed by external scintiscanning of the upper abdomen with a gamma camera. When the gallbladder is imaged by the isotope, contraction and emptying of the organ is achieved either with a milk meal (milk–EHIDA test) or by intravenous cholecystokinin (EHIDA–CCK test). The amount of enterogastric reflux is calculated as a percentage of the total abdominal radioactivity (Figure 23.19).

The symptoms of reflux gastritis and bile vomiting may be improved by the administration of bile salt-binding agents (cholestyramine, aluminium hydroxide, charcoal). However, conservative management along these lines often fails, when remedial surgical intervention becomes necessary. Prolonged enterogastric reflux can result in atrophic gastritis and intestinal metaplasia. It has been incriminated as a factor in the development of carcinoma of the stomach after gastric surgery, although evidence for this hypothesis remains lacking.

Extrinsic loop obstruction

This rare complication occurs after truncal vagotomy and gastrojejunostomy and usually affects the afferent loop. The predisposing factors to the development of afferent loop obstruction include antecolic anastomosis of long loops (exceeding 20 cm). The causes of extrinsic loop obstruction are:

- internal herniation
- kinking of the anastomosis
- adhesions
- volvulus
- stenosis
- jejunogastric intussusception
- development of carcinoma of the gastric remnant.

Obstruction of afferent or efferent loops is usually chronic and intermittent but may be acute.

The symptoms of chronic afferent loop obstruction include fullness, cramp-like pain and nausea within 1 hour of eating. The attack culminates in vomiting of copious amounts of bile-stained fluid which relieves the symptoms. The presentation of acute afferent loop obstruction is with severe colicky abdominal pain, nausea and vomiting, which is characteristically free of bile. Abdominal tenderness is present. The condition may be complicated by the development of acute pancreatitis, jaundice and necrosis with perforation.

Acute jejunogastric intussusception is a serious condition characterized by severe epigastric pain, vomiting, haematemesis, a palpable abdominal mass with high small bowel obstruction. Urgent surgical intervention is required because of the risk of strangulation and gangrene. The condition may be diagnosed preoperatively by a plain abdominal film which shows a soft-tissue epigastric mass surrounded by the gastric air bubble. Alternatively, emergency barium meal or endoscopy will establish the diagnosis.

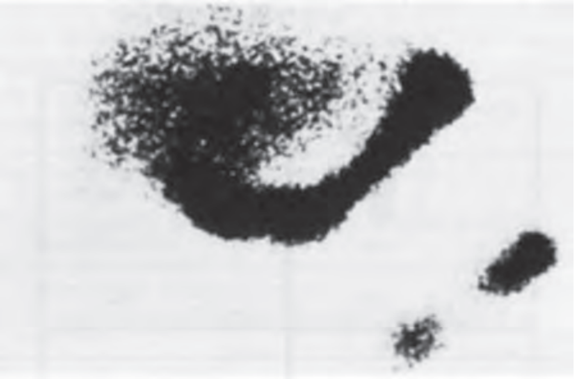


Figure 23.19 The milk–EHIDA test for enterogastric reflux. The patient experienced severe bile vomiting after vagotomy and pyloroplasty. Reflux of isotope (bile) into the stomach occurs in the fasting state but is considerably enhanced after the administration of milk (as a source of fatty meal).

Gastro-oesophageal reflux and oesophagitis

The situation regarding gastro-oesophageal reflux and surgery for duodenal ulcer is both confusing and controversial. In the first instance gastro-oesophageal reflux often accompanies duodenal ulcer and oesophagitis may, therefore, be present preoperatively. Transient dysphagia may occur after any type of vagotomy and this has been attributed to oedema of the lower abdominal oesophagus. It is now established that vagotomy itself does not affect the oesophageal high-pressure zone, but damage to the oesophageal attachments, particularly the phreno-oesophageal membrane during the mobilization of the oesophagus, may cause cardio-oesophageal incompetence. If this becomes subsequently associated with the development of enterogastric reflux of bile and pancreatic juice (neutral or alkaline reflux) a severe form of erosive oesophagitis develops. It is now documented that the eradication of *H. pylori* in patients with duodenal ulcer aggravates any underlying cardio-oesophageal incompetence and has been suggested as one of the factors accounting for the rising incidence of junctional cancers during the past 20 years.

Diarrhoea

The reported incidence of this complication varies widely, largely due to varying definitions. Three patterns of diarrhoea are encountered after gastric surgery:

- frequent loose motions
- intermittent episodes of short-lived diarrhoea
- severe intractable explosive diarrhoea.

Severe explosive diarrhoea is a serious, but rare disability, being encountered in 2% of patients after truncal vagotomy with drainage. It is often accompanied by dumping symptoms and is precipitated by food. Severe intractable diarrhoea is characterized by extreme urgency and often causes incontinence during an acute attack. Also the fluid motion is extremely foul smelling and this is often complained of by the partner and close family members.

Although often associated with rapid gastric emptying, the exact mechanism is unknown. Malabsorption of bile salts and/or fatty acids consequent on the intestinal vagal denervation has been implicated but never confirmed. The small bowel transit is markedly exaggerated in all patients.

A full malabsorption survey is necessary in all patients with severe diarrhoea as in a few patients this disability is secondary to a previously undiagnosed gastrointestinal disease (e.g. adult coeliac) or bacterial overgrowth consequent on a blind loop. Medical management is with a low animal fat diet, intestinal sedatives (codeine phosphate, loperamide) and bile salt-binding agents such as cholestyramine. Although temporary improvement can be obtained in this way, long-term benefit is rarely obtained with conservative measures and the patient either accepts and lives with the disability or asks for remedial surgical treatment.

In the past 10 years, cases of severe explosive diarrhoea have been reported after laparoscopic fundoplication especially of the anterior (Watson) type. The incidence after this operation is about 2%. The presumed aetiology is damage to the vagal nerve trunks. These patients should be treated conservatively for 2 years as a considerable number resolve by this time. Remedial surgery may be considered if the condition persists beyond 2 years.

Small stomach syndrome

This term is sometimes used for the early satiety complained of by many patients after vagotomy, which causes loss of the receptive relaxation of the stomach during eating. It is however, best reserved for those unfortunate patients, usually females, who cannot eat after gastrectomy and whose quality of life is miserable and whose nutrition can only be maintained by enteral or parenteral feeding. In essence, these patients, because of pain and myriad other symptoms, are unable to eat and in essence starve themselves. Thus the condition inevitably leads to protein calorie malnutrition and cachexia similar to that encountered in advanced malignancy. In the author's experience, small stomach syndrome is refractory to conservative management, although in some patients nutrition can be maintained by elemental diets administered via a Clinifeed tube and an IVAC pump or via a feeding jejunostomy in those who cannot tolerate nasoenteral feeding. Although many can be trained to use this in their homes and maintain a reasonable nutritional state in this way, their quality of life is extremely poor and it is no wonder that some commit suicide. The alternative to feeding jejunostomy is home hyperalimentation with intravenous nutrition. Aside from cost, this is best avoided in these patients because of the inordinately high complication rates associated with the indwelling intravenous line. If the patient is fit and middle aged, surgical intervention designed to reconstruct a gastric reservoir and restore duodenal continuity provides the best chance for improvement and ability to maintain nutrition with an oral diet although supplementation is often needed.

Other complications

These include the formation of gallstones and bezoars and the development of gastric carcinoma. Vagotomy causes dilatation of the gallbladder. However, although there are a number of reports indicating an increased risk of gallstone formation after both vagotomy and partial gastrectomy, there is no firm evidence that gastric surgery predisposes to cholelithiasis.

The factors implicated in the formation of bezoars after gastric surgery include hypoacidity, impaired proteolytic activity, inadequate mastication and loss of the antral pump. The majority of bezoars, which develop after gastric surgery, consist of undigested vegetable/fruit fibre debris (notably orange pith). Bezoars can cause chronic symptoms such as nausea, vomiting, abdominal discomfort, halitosis and early satiety. They can also lead to serious complications, e.g. small bowel obstruction, severe gastritis and ulceration, bleeding, perforation and malnutrition. Treatment is initially conservative by enzyme (cellulose) digestion or endoscopic fragmentation/removal. Surgical intervention is undertaken if medical/endoscopic therapy fails or because of development of a complication.

There is good evidence that previous gastric surgery (partial gastrectomy, gastrojejunostomy) predisposes to the development of gastric carcinoma in the stomach remnant. Vagotomy does not appear to be implicated in this complication. Although reflux gastritis with the development of intestinal metaplasia, particularly of the type III variety, and bacterial overgrowth with the formation of nitrosamines in the hypochlorhydric gastric remnant have been implicated, the exact mechanism for the development of invasive carcinoma remains unknown. Also it seems likely that many of these patients had *H. pylori* gastritis that was not recognized at the time. This is now thought to have contributed to the development of gastric carcinoma. There is a long latent period of 15–20 years but the risk, though definite, is small.

Remedial surgical treatment for postgastric surgery syndromes

General considerations

As the majority of symptoms improve with time, initial management should always be conservative and surgical treatment should not be considered before 18 months to 2 years of presentation and with no improvement on conservative management. Only patients with severe symptoms which persist beyond this time despite conservative management should be considered as candidates for remedial surgery, and, ideally, the patient should request the intervention. All these patients need to be investigated and assessed (nutritionally and fitness for surgery). In addition, a detailed history of the symptoms during 'bad days' is essential to establish the *dominant symptom/disability* – what troubles the patient most and would improve his or her quality of life if treatment were to be successful. Remedial surgery should be directed to amelioration of this dominant symptom with the proviso, which must be conveyed to the patient, that while the intended surgery may impart considerable benefit, a totally symptom-free outcome is rarely, if ever, obtained. Despite the classical descriptions of separate syndromes, most patients cannot be pigeonholed and, instead, have a mixture of symptoms, but a careful history will always determine the dominant symptom/disability and remedial surgery should be directed exclusively to correction of this symptom if the patient is to benefit from remedial surgery.

Dominant dumping symptoms

The easiest patients to manage surgically are those who experience severe dumping after truncal vagotomy with drainage. In these patients, the remedial surgery consists of either take down of the gastrojejunostomy or pyloric reconstruction (Figure 23.20) depending on the type of drainage used. Surprisingly, symptomatic gastric retention is rarely encountered, although postprandial fullness is common and the patient should be advised to avoid heavy bulky meals, lying down after meals and to eat the evening meal at least 3 hours before retiring to bed. In practice, prokinetic drugs do not help these patients.

For patients with dominant dumping symptoms after partial gastrectomy or vagotomy with antrectomy, the more technically demanding isoperistaltic jejunal interposition (10–15 cm) between the gastric remnant and the duodenum acts as an effective brake to slow the gastric emptying (Figure 23.21).

Bile vomiting as the dominant symptom

The commonest cause of this is enterogastric reflux. Excellent results are obtained in patients with vagotomy and gastrojejunostomy by take down of the latter. In contrast, pyloric reconstruction for bilious vomiting in these patients gives unsatisfactory results and should not be attempted. These patients are best served by an antrectomy and an isoperistaltic jejunal interposition (10–15 cm) between the proximal stomach and the duodenum. Severe bile vomiting due to enterogastric reflux after partial Billroth I gastrectomy may be treated by Roux-en-Y diversion. Although very effective in abolishing the bile vomiting, this procedure may lead to bacterial overgrowth, but this is not usually a major problem. An alternative approach is reconstruction with an isoperistaltic jejunal loop between the gastric remnant and the duodenum.

Dominant symptoms of the small stomach syndrome

This is the most difficult symptom complex to treat. These patients are usually grossly malnourished and require a period of parenteral nutrition before the remedial surgical treatment. The best procedure consists of completion gastrectomy and the creation of a jejunal reservoir with an isoperistaltic conduit between the oesophagus and the duodenum (Figure 23.22).

In these patients, the gastric stump is invariably inflamed and immotile and is therefore excised prior to commencement of the reconstruction. An improved outcome following this major reconstruction is reflected by the ability of the patient to sustain weight by a semisolid low-roughage diet and is encountered in 50–60% of patients. The problem has been the prediction of patients who are likely to benefit from such extensive remedial surgery. Obviously, psychological factors are important in this respect and the patient requires to be well motivated to overcome the disability. Too much stress cannot, however, be based on psychological assessment as the psychological state of the patient may be the result of the disability. Preoperative weight gain on enteral feeding is reported to be a good indicator of a positive clinical outcome after remedial surgery for the small stomach syndrome (Figure 23.23). In patients who do not improve after completion of gastrectomy and creation of a gastric reservoir, or those who are unfit for major surgery, nutrition can be maintained by surgically constructed feeding jejunostomy. This is physiologically better, safer and more cost effective than permanent home hyperalimentation with parenteral nutrition.

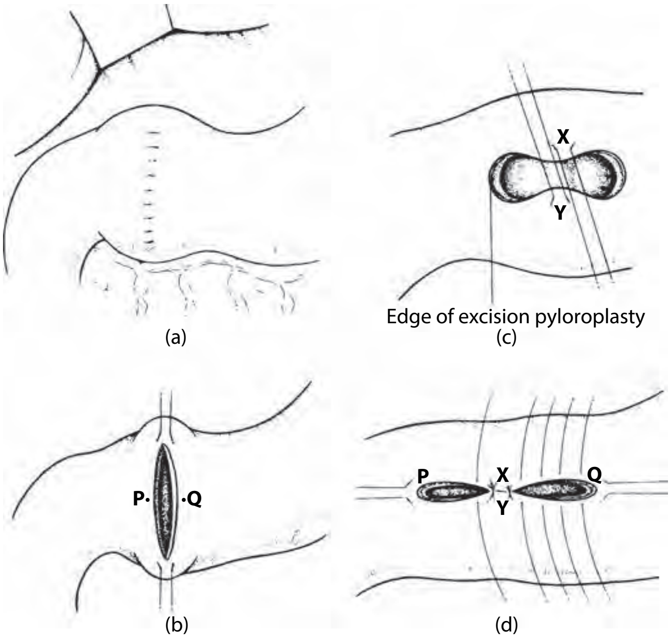


Figure 23.20 Pyloric reconstruction for dumping after Heineke–Mikulicz pyloropasty: (a) complete exposure of the pyloroplasty by freeing of adhesions; (b) insertion of stay sutures beyond both ends of the pyloroplasty scar which is carefully incised along its length; (c) precise restoration of the normal alignment of the pyloric muscular ring; (d) the realigned antropyloric segment is closed in a single layer with fine (3/0) carefully placed interrupted sutures.

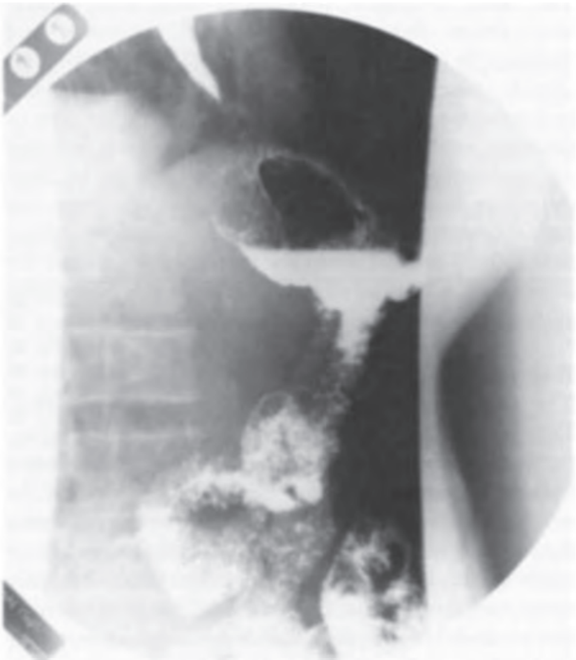


Figure 23.21 Contrast radiology of isoperistaltic jejunal interposition between the gastric remnant and the duodenum.

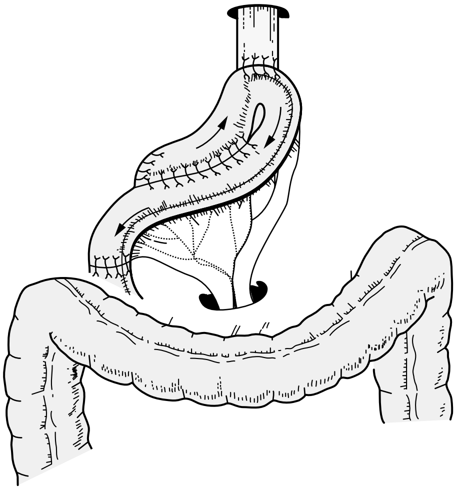


Figure 23.22 Jejunal pouch reconstruction with an isoperistaltic conduit for the small stomach syndrome. The useless gastric remnant is first removed before this reconstruction.

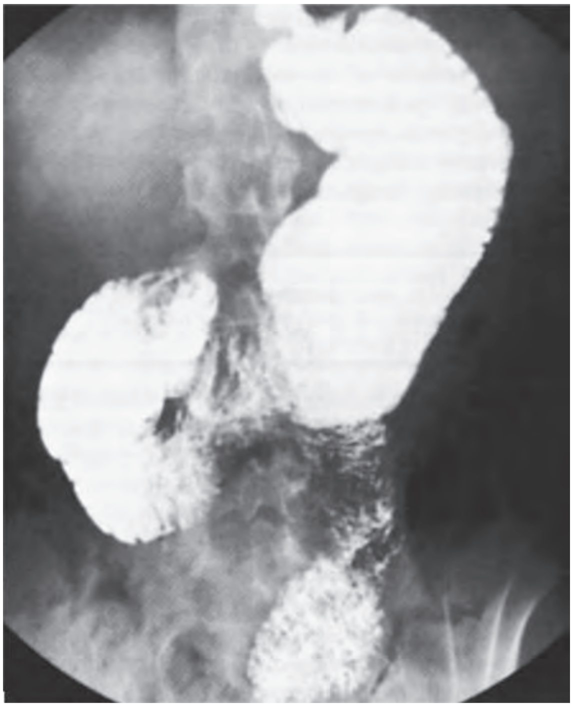


Figure 23.23 Barium contrast series in a patient with completion gastrectomy and reconstruction with a jejunal reservoir with an isoperistaltic conduit. The operation was performed for severe symptoms (small stomach syndrome) with good symptomatic result and weight gain on a semifluid oral diet.

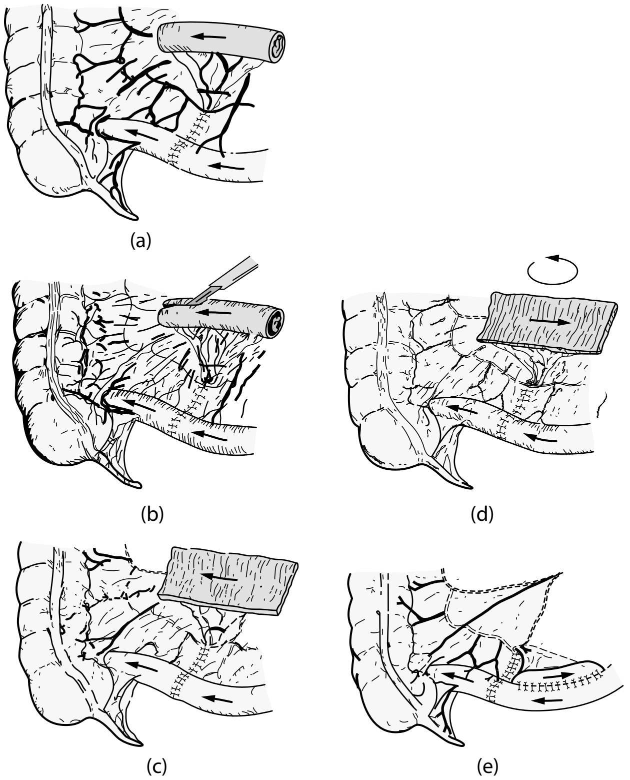


Figure 23.24 Reversed ileal onlay graft for severe explosive diarrhoea: (a) a 10–12 cm segment of ileum is isolated on an intact vascular pedicle some 30 cm proximal to the caecum; (b) continuity of the small bowel is restored by an enteroenteric anastomosis; (c) the isolated segment is split longitudinally along its antimesenteric border; (d) reversal of the flap; (e) suture of the reversed graft as an onlay to the adjacent ileum following an appropriate enterotomy, thereby creating a passive non-propulsive segment. (From Cuschieri, *Br J Surg* 1986;**73**:981, with permission).

Severe explosive diarrhoea

Although the use of reversed jejunal segments has been advocated, the outcome of these operations is poor due to the development of episodes of postprandial colic, intestinal obstruction, distension and bacterial overgrowth. The best results are obtained by the distal onlay ileal graft procedure, which is designed to create a passive non-propulsive segment of the small intestine some 30–60 cm from the ileocaecal junction (Figure 23.24). Some 50% of patients obtain significant benefit following this operation if they avoid fatty/dairy products. Some patients complain of postprandial colic after this operation with some abdominal distension. Revision (reduction of the length of the distal onlay graft) may be necessary if these symptoms persist.

Remedial surgery fails in half the patients and, in these, a temporary loop ileostomy can be offered. If the patient finds this acceptable after a trial period of 3 months, the ileostomy is made permanent, otherwise it is reversed.

Gastritis

There is considerable confusion regarding the nomenclature of inflammatory conditions of the gastric mucosa (gastritis). Clinical designations do not always conform with the specific histological patterns and, in some instances, vascular changes are present with minimal cellular inflammatory response (Box 23.5). The term *gastropathy* as opposed to gastritis is more appropriate to the latter conditions, typified by alcoholic gastritis.

In the past, there have been several clinical and histological classifications of gastritis producing a rather confusing picture to the clinician. In 1990, the Sydney system was introduced as a

BOX 23.5 Classification of gastritides

Common
- Non-atrophic gastritis (*H. pylori*)
- Atrophic gastritis type A (autoimmune)
- Atrophic gastritis type B (associated with *H. pylori* infection)
- Chemical/reactive gastritis
- Stress gastritis (seriously ill patients)

Rare
- Lymphocytic gastritis
- Granulomatous gastritis (tuberculosis, Crohn disease)
- Gastritic cystic polyposa
- Gastritis in immunosuppressed patients
- Eosinophilic gastritis
- Radiation gastritis
- Suppurative phlegmonous gastritis
- Emphysematous gastritis

basis for classification and this is now widely accepted and was updated at the international workshop in Houston in 1966; hence it is referred to as the Sydney–Houston classification. The system is based on the histological examination of at least four biopsies, two from the middle body and two from the antrum at least 2.0 cm proximal to the pylorus (anterior and posterior walls). The Houston update also recommends that additional biopsies from the lesser curvature and the angular notch are necessary especially for the early diagnosis of atrophic gastritis and intestinal metaplasia. The main change in the classification of gastritis outlined by the Houston update is the designation of two broad categories: *atrophic* and *non-atrophic*, although there is some disagreement on the validity of this subdivision as a mixed picture may be encountered (gland atrophy and metaplasia).

The new classification is based on five histological variables that are graded separately by the histopathologist as mild, moderate or severe:

- chronic inflammation – mononuclear infiltrates
- activity – acute polymorphonuclear infiltrates
- atrophy – loss of normal glands
- intestinal metaplasia
- extent of colonization of biopsies by *H. pylori* in non-metaplastic epithelium.

Extensive lymphoid follicle formation is indicative of *H. pylori* infection and may lead to the development of MALT lymphomas In general, the grade (mild to severe) of the colonization correlates with the severity of the mononuclear inflammation, the activity (polymorphonuclear inflammation), the mucus depletion, lymphoid follicle formation and the degenerative changes. In practice, the system results in the following categorization of the gastric mucosa:

- normal
- non-atrophic *H. pylori* gastritis
- atrophic gastritis – *H. pylori* positive (type B) or negative (type A)
- special forms of gastritis.

Patients with a histologically normal stomach by the Sydney–Houston classification have a normal functional mucosa with respect to the output of acid, pepsinogens, IF and peptide hormones, and these individuals have a low risk of peptic ulcer and gastric carcinoma. Ulcers when they occur are due to drug ingestion such as NSAIDs or steroids and rarely to Crohn disease or rarer still to the sporadic or hereditary forms of Zollinger–Ellison syndrome.

H. pylori and much more rarely *H. heilmannii* cause both non-atrophic and atrophic gastritis. *H. pylori* non-atrophic gastritis involves mainly the antrum and the duodenal bulb (bulbitis) and is strongly associated with the development of peptic ulcers with a cumulative risk of symptomatic peptic ulcer disease of 30% over a period of 10 years, the infection having been acquired in childhood.

Atrophic gastritis

Atrophic gastritis can occur in the absence of *H. pylori* infection (type A) or be caused by infection with this organism (type B). In both types, the risk of peptic ulceration, particularly duodenal, is low. The secretory function of the gastric mucosa becomes progressively impaired and if the antral mucosa is severely involved, there is a reduction/loss of the gastrin-secreting G-cells and the somatostatin-secreting D-cells.

Type A atrophic gastritis (autoimmune gastritis)

This affects mainly the body and fundus of the stomach and has an autoimmune aetiology with the presence of circulating autoantibodies. These are directed against at least three antigens: IF, parietal cell cytoplasmic (microsomal–canalicular) and plasma membrane antigens. Two types of IF antibodies are present (types I and II). The type I IF antibody is the more important as it blocks the IF–cobalamin binding site, thus preventing the uptake of vitamin B_{12}. As part of the cell-mediated immunity, T-cell lymphocytes infiltrate the gastric mucosal cells and contribute to epithelial cell destruction and gastric atrophy of the parietal and chief cells, which leads to achlorhydria and loss of secretion of IF with malabsorption of B_{12}, and, eventually, pernicious anaemia in some patients. As the antrum is relatively spared, the serum gastrin is elevated and this chronic hypergastrinaemia may lead to the development of endocrine tumours. There is an increased risk of gastric cancer, which is usually of the diffuse type.

Autoimmune gastritis is a relatively rare disease, most frequently encountered in individuals of northern European descent and blacks. The prevalence of pernicious anaemia is estimated at 127 cases per 100 000 members in the UK, Denmark and Sweden. In these countries pernicious anaemia is usually diagnosed in patients aged 60 years. Aside from the haematological manifestations (megaloblastic anaemia and rarely purpura), the patients exhibit symptoms of anaemia (weakness, light-headedness, vertigo and tinnitus, palpitations, angina) and may develop congestive heart failure. Some patients complain of a sore tongue. Anorexia with moderate weight loss with diarrhoea may also be present. Delayed diagnosis and treatment may result in neurological manifestations resulting from demyelination of peripheral nerves, posterior and lateral columns of the spinal cord and cerebrum (numbness and paraesthesias in the extremities, weakness, and ataxia). Sphincter disturbances may occur. Mental functional disturbances vary from mild irritability to severe dementia or psychosis. Patients with pernicious anaemia have an increased frequency of gastric polyps and gastric carcinoids (2.9-fold increase) together with an increased risk of invasive gastric cancer.

Type B atrophic gastritis

This is the commonest form of chronic gastritis and hence its clinical importance. It is the result of infection by *H. pylori* and eradication of the infection is followed by documented histological improvement although relapse may occur. *H. pylori* causes mucosal tissue damage with a mixed histological picture of both acute and chronic gastritis. There is T- and -cell lymphocyte infiltration as part of the host's chronic response to the infection. The acute response to the infection consists of polymorphonuclear infiltration of the lamina propria and gastric epithelium, often seen on histological sections to be phagocytosing the bacteria. The interaction of *H. pylori* with the gastric mucosa leads to the release of the proinflammatory cytokine interleukin (IL)-8, which is responsible for the recruitment of polymorphonuclear lymphocytes. Various other cytokines, including tumour necrosis factor (TNF)- and multiple interleukins (e.g. IL-6, IL-8, IL-10), are also released in the gastric mucosa of infected patients. Leukotriene levels are also secreted. In particular, leukotriene B4 (synthesized by host neutrophils) is paradoxically cytotoxic to gastric epithelium. In time, this mixed inflammatory response leads to structural and functional changes in the stomach, the outcome of which will depend on the areas involved: in gastric body and fundus, parietal and chief cells are destroyed (hypochlorhydria), whereas antral inflammation affects G- and D-cell function, gastrin secretion is abnormal in infected individuals and usually consists of an exaggerated meal-stimulated release of gastrin.

Individuals infected with *H. pylori* strains that secrete the vacuolating toxin A (vacA) are more likely to develop peptic ulcers. Other virulence factors are encoded by the *H. pylori* pathogenicity PAI, which contains the sequence for several genes and encodes the *CAGA* gene. Strains producing the CagA protein (CagA$^+$) are associated with a greater risk of development of both gastric carcinoma and *H. pylori*-associated chronic gastritis progresses along two patterns according to site of predominant involvement: *antral predominant gastritis* (APG) and *multifocal atrophic gastritis* (MFAG). APG is characterized by inflammation that is largely limited to the antrum. These individuals develop peptic ulcers, duodenal ulcer if the parietal cell mass is large usually demonstrate this pattern of gastritis. MFAG is characterized by involvement of the whole stomach with progressive development of gastric atrophy (loss of the gastric glands) and partial replacement by intestinal metaplasia. These patients are at risk of developing both gastric ulcers and gastric cancer. Some individuals also develop gastric MALT lymphomas. The normal healthy stomach lacks organized lymphoid tissue, but develops lymphoid tissue infiltrates in response to the infection with *H. pylori*. If the infection is not eradicated, the persistence of MALT aggregates may eventually progress to the development of both low- and high-grade MALT lymphomas (monoclonal proliferations of neoplastic -cells).

Gastric carcinomas develop in a small proportion of individuals with extensive atrophy and intestinal metaplasia. Persistence of the infection over a long period of time is thought to induce mutations in the genome of the gastric epithelial cells, leading to an increased risk of malignant transformation to invasive gastric adenocarcinoma. These mutations in the gastric epithelium are secondary to oxidative DNA damage associated with chronic inflammatory by-products caused by the chronic bacterial infection.

Special forms of gastritis

Reactive/erosive/chemical gastritis

This is sometimes referred to as *gastropathy* as the inflammatory component is not marked. It results from gastric mucosal damage by both exogenous and endogenous irritant chemicals. Histologically, there is foveolar hyperplasia, severe congestion, oedema and fibrosis of the lamina propria but a paucity of inflammatory cells. Reactive gastritis is commonly caused by drugs, e.g. NSAIDs (present in 25–45% of NSAID users) and alcohol. The usual locations of drug-induced chemical gastropathy are the antral and prepyloric regions. The lesions are produced by blockade of the cyclo-oxygenase pathway with reduction of the cytoprotective gastric prostaglandins. Thus NSAIDs that are prostaglandin sparing (non-acetylated salicylates such as carpofen, nabumetone, etc.) and low-dose steroids are less likely to produce erosive gastropathy. The alcohol-induced mucosal damage affects in addition the mucosal microvessels which undergo necrosis with resulting haemorrhage and thrombus formation. The role of leukotrienes in this pathological process is debatable.

Chemical gastropathy also develops as a result of enterogastric reflux usually in patients after partial gastrectomy. Several bile constituents are responsible for the damage: 5% lysolecithin and bile acids disrupt the gastric mucous barrier and thus cause back diffusion of positive hydrogen ions and cellular injury. Pancreatic juice also damages the gastric epithelium by virtue of its enzyme content. Other causes of haemorrhagic gastropathy with erosions include cor pulmonale, severe infections such as pneumonia, cirrhosis and blood disorders.

Lymphocytic gastritis

This rare form of gastritis accounts for 1–5% of cases of patients presenting with dyspepsia but is more commonly encountered among patients with coeliac disease in whom it is present in 15–45%. It is thought to be the result of an abnormal immunological reaction to unidentified luminal antigens. The condition may be intermittent. Histologically, it is characterized by infiltration of the surface gastric foveolar epithelium associated with chronic infiltrates in the lamina propria by T-lymphocytes similar to those encountered in coeliac disease (intraepithelial lymphocytes greater than 25/100 epithelial cells). Some cases are associated with infection by *H. pylori*. Antibody titres are certainly found in some patients with lymphocytic gastritis, and the gastritis resolves after *H. pylori* eradication. The majority of patients with lymphocytic gastritis are, however, serologically negative for *H. pylori*. Few cases of lymphocytic gastritis have been reported secondary to intolerance to drugs such as ticlopidine.

Although some patients do not exhibit any specific endoscopic features, in the majority a distinctive appearance consisting of nodularity, erosions and enlarged mucosal folds (referred to as varioliform gastritis) is observed. Lymphocytic gastritis is most commonly encountered in middle age. Lymphocytic gastritis may be associated with chronic *H. pylori* infection, gluten-sensitive enteropathy, Ménétrier disease and has been reported in both MALT lymphoma and gastric carcinoma.

Gastritis cystic polyposa

This is a rare late complication of gastric surgery (1–25 years), although instances without a history of previous gastric operations have been reported. The patients present with abdominal pain, nausea and vomiting or gastrointestinal bleeding. Endoscopically, a hypertrophic gastritis is present and this may be mistaken for carcinoma on the macroscopic appearance.

Gastritis in immunosuppressed patients

These patients are prone to infection by various viruses including cytomegalovirus, herpes simplex and bacteria (*Mycobacterium avium-intracellulare*) and parasites. In infections by cytomegalovirus, the histology shows the typical intranuclear eosinophilic inclusions with a patchy mild inflammatory infiltrate. Severe necrosis may result in ulceration. Infections with herpes simplex are diagnosed by basophilic intranuclear inclusions in epithelial cells. Infections by *M. avium-intracellulare* cause diffuse infiltration of the lamina propria by histiocytes.

In patients with AIDS vomiting due to gastric outlet obstruction from gross oedema of the pyloric ring is caused by infection with *Cryptosporidium*. In these patients, cryptosporidial oocysts can be recovered from the stools. An interesting confirmed observation is the low prevalence of *H. pylori* infection in HIV-positive patients with low CD4 counts (<200) compared to HIV-negative patients. This low prevalence is also accompanied by a reduced incidence of peptic ulcers. Gastric toxoplasmosis due to infection with *Toxoplasma gondii* is rare in AIDS patients, but when it occurs it causes abdominal pain. Diffuse thickening of the gastric folds is seen at endoscopy. Gastric mucosal biopsy confirms necrosis and intracellular trophozoites in gastric epithelial, smooth muscle and endothelial cells.

Eosinophilic gastritis

This occurs as part of eosinophilic gastroenteropathy in children and adults below 50 years and has an allergic basis (allergy to milk or soya protein especially in children). Some cases of eosinophilic gastritis are caused by parasitic infections, e.g. *Eustoma rotundatum* and anisakiasis. Eosinophilic gastroenteritis can also be found in patients with connective tissue disorders (polyarteritis nodosa, scleroderma, polymyositis and dermatomyositis).

The pyloric regions and the adjacent duodenum become diffusely thickened due to oedema of the submucosal and muscle layers, which are also infiltrated with eosinophils and occasional giant cells. The gastric antrum is the area of the stomach most severely affected. Histological examination of biopsies shows extensive eosinophilic infiltrates with oedema and lymphangiectasia. The elevation of the serum IgE correlates with the severity of the disease. In the majority of patients there is a peripheral eosinophilia. Cases arising as a complication of polyarteritis nodosa tend to be severe and life threatening. The clinical features of eosinophilic gastritis include symptoms of delayed gastric emptying, e.g. early satiety, nausea and vomiting and gastrointestinal bleeding. Patients with established food allergy often respond to removal of these items from the diet. Otherwise, the treatment is with sodium cromoglycate and/or prednisolone.

Granulomatous gastritis

Granulomatous gastritis may be infectious or non-infectious. Infectious granulomatous gastritis is rare and usually secondary to active pulmonary diseases, although caseating granulomas may be found in the absence of lung disease in malnourished and immunosuppressed patients. Tuberculosis infection causes multiple ragged ulcers and discrete tubercles may be visualised at endoscopy. Serosal inflammation is common and there is marked locoregional lymphadenopathy. Treatment is with antituberculous chemotherapy. Fungal infections can also cause caseating granulomas and necrosis, also in immunosuppressed patients.

Non-infectious granulomatous gastritis is usually caused by Crohn disease when it presents with gastric pain, nausea and vomiting. The reported estimates of Crohn's gastritis in patients with established intestinal Crohn disease varies from 6% to 24%. However, almost invariably, patients with gastric Crohn disease have associated small bowel disease which dominates the clinical presentation. Non-caseating granulomas are present in one-third of cases and involvement of the duodenum is common. In addition there is focal chronic active ulceration with erosion of the epithelium in the absence of *H. pylori* infection, which should suggest the diagnosis.

Sarcoidosis of the stomach is another cause of chronic non-infectious gastritis but is usually accompanied by granulomatous inflammation in other locations (lungs, hilar nodes or salivary glands). A few patients with gastric sarcoid involvement in the stomach have no gastric symptoms but others present with gastric ulcers, haemorrhage, and pyloric stenosis causing gastric outlet obstruction. Rarely, the granulomatous gastritis is idiopathic (diagnosis made after exclusion of all known causes). These patients are usually older than 40 years, and present with epigastric pain, weight loss and vomiting secondary to pyloric obstruction.

Radiation gastritis

Although small doses of radiation (up to 1500 R) cause reversible mucosal damage, higher radiation doses induce irreversible damage with atrophy and ischaemic-related ulceration from submucosal endarteritis. The severely damaged mucosa causes repeated episodes of bleeding and may require resection.

Suppurative (phlegmonous) gastritis

This is a rare and often fatal bacterial infection producing a cellulitis of the stomach wall. Haemolytic streptococci are the commonest infecting organisms. The condition usually complicates a pre-existing gastric lesion and is more commonly encountered in elderly and alcoholic patients. The presentation is usually one of severe progressive peritonitis. Treatment is with gastric resection and systemic antibiotics. At operation, the stomach exhibits a dusky discoloration and its serosal surface is covered with a fibrinous exudate. Necrotizing gastritis is an especially severe variant which results in overt infective gangrene. It is caused by a mixed infection with fusiform and spirochaete bacteria from the mouth.

Emphysematous gastropathy

Air-filled cysts in the gastric wall are rare and usually accompanied by *pneumatosis cystoides intestinalis*. Gas cysts in the wall of the stomach can also occur in association with pyloric obstruction and chronic obstructive airway disease with emphysema. Gas can also be introduced in the wall of the stomach following incomplete injuries during endoscopy.

Clinical management of gastritis

It is important to stress that, irrespective of symptoms and presentation, the specific diagnosis in the individual patient rests on endoscopy and biopsy, which is mandatory in all patients. The management then depends on the severity and effects of the gastritis (e.g. bleeding, acute or chronic symptoms), the removal of the underlying cause when applicable, or specific therapy (medical or surgical) when indicated. When the chronic gastritis is accompanied by dysplasia, surveillance by endoscopy is indicated and if the dysplasia becomes severe, then gastric resection is indicated.

Gastric tumours

Gastric adenocarcinoma

Incidence

Worldwide there is an estimated 930 000 new diagnoses and 700 000 deaths from gastric cancer. Overall incidence rates are 22 and 10.3 per 100 000 per annum in males and females respectively with the corresponding mortality rate being 14.3 per 100 000. Gastric cancer is the third most common cause of male cancer and the second most common cause of male cancer death. It is the fifth most common cause of female cancer and the fourth most common cause of female cancer death. Incidence rates in males are approximately double those in females and both sexes are strongly related to age. Most patients are between 60 and 80 years old at diagnosis.

There is a wide geographical variation in gastric cancer with the highest incidence in East Asia, South America and Eastern Europe and the lowest incidence seen in North America, Western Europe and Africa. Explanations of the geographical variation have been sought. High intake of traditional salt-preserved food and salt and low consumption of fresh fruit and vegetables are associated with a high risk of gastric cancer. In support of this observation, the gastric cancer incidence in migrants from low-incidence countries increased from a low rate in first-generation migrants to high incidence of their host country in the second generation. Also, *H. pylori* is a recognized risk in the development of gastric cancer. However, not all populations with high rates of *H. pylori* infection, such as Africa and South Asia, have a raised incidence of gastric cancer. Smoking is another environmental risk factor for gastric cancer.

The incidence of gastric cancer is declining. Even assuming the current decline, the incidence of gastric cancer will continue: the predicted growth in the world population combined with the increased longevity will most likely result in a net increase in the overall number of gastric cancers being diagnosed. Nevertheless, in the Western population there is an increase in adenocarcinoma of the gastric cardia, despite the decline in the incidence of gastric cancer.

Aetiology

Hereditary diffuse gastric cancer accounts for 1–3% of gastric cancer. In a third of familial gastric cancers, germline mutation in allele 1 of the E-cadherin gene (*CDH1*) is identified. Activation of the second allele occurs by either mutation or hypermethylation. Additional genotype changes lead to early onset of diffuse gastric cancer. The estimated life risk of gastric cancer in carriers of the *CDH1* mutation is 67% in men and 83% in women. Other hereditary syndromes that raise gastric cancer risk include Lynch syndrome (mutation in one of the mismatch repair gene) and Peutz–Jeghers syndrome (*SK11* mutation).

One of the most important aetiological factors associated with both intestinal and diffuse gastric cancer is *H. pylori*. Exposure to *H. pylori* results in inflammatory action and the production of reactive oxygen species and nitrous oxide which in turn deaminates DNA causing mutation. People who are positive to *H. pylori* have at least sixfold greater risk of developing gastric adenocarcinoma than those without infection. Long-term *H. pylori* infection may lead to atrophic gastritis and intestinal metaplasia with increased relative risk for developing gastric cancer ranging from 1.7 in moderate atrophy, 4.9 in severe atrophy to 6.4 in intestinal metaplasia. However, despite the high incidence of *H. pylori* infection in Africa and South Asia, the incidence of gastric cancer in those areas is much lower than in other countries. The incidence of gastric cancer tends to decrease from north to south in East Asia. Such geographical differences in the pathology can be explained at least in part by the presence of different types of *H. pylori* virulence factors, especially CagA, VacA and the right end of the Cag PAI.

Many studies showed that a diet high in fruit and vegetables reduces the risk of developing gastric cancer. While there are many potential mechanisms by which a diet rich in fruit and vegetables may be protective against gastric cancer, the active component is not clear. One possibility is related to the antioxidant capacity of fruit and vegetables. However, recent analysis did not find evidence that antioxidant supplements prevent gastrointestinal cancer in general or stomach cancer in particular.

A possible increase in risk of gastric cancer associated with alcohol intake has been suggested by meta-analysis of cohort and case–control studies of gastric cancer. However, on the basis of significant heterogenicity observed between studies, it is not possible to infer causality from these data.

Evidence that salt intake is inversely associated with stomach cancer has accumulated over many decades. There are many problems evaluating such studies because of the methods of measuring salt intake, the irritation effect of salt on the stomach lining and the development of atrophic gastritis is compounded by the effect of *H. pylori* infection, which is difficult to adjust for these studies. Also, the consumption of salty food commonly correlates with the intake of fresh vegetables and as a method of preservation and storage, which are difficult to control for those studies. Smoking is another aetiological factor. The results from both cohort and case–control studies are consistent with a causal role of tobacco smoking in the development of gastric cancer. There is a higher risk of gastric cancer among smokers than non-smokers in the order of 1.5–1.6.

With regard to PPIs, there is now sufficient experience to allow reasonable assessment of their safety in terms of cancer development. Long-term PPI use is associated with an increase in gastric inflammation and development of atrophy among those with active *H. pylori* infections. The actual risk is unknown but is certainly low. However, it can be markedly reduced or eliminated by *H. pylori* eradication leading to the recommendation that patients considered for long-term PPI therapy be tested for *H. pylori* infection and if present it should be eradicated. Oxyntic cell hyperplasia, glandular dilatations, and fundic gland polyps may develop in *H. pylori*-uninfected patients, but these changes are believed to be reversible and without significant cancer risk.

Histopathology

Gastric adenocarcinoma has two main types: the intestinal type and diffuse cancer. The *intestinal type* has well-defined ductal structures surrounded by desmoplastic stroma reaction containing different amounts of mixed inflammatory infiltrates. Tumour cells are large and nuclei are polymorphic and have a course chromaffin pattern. Mitotic figures are easily detected. Intestinal type carcinoma is usually well to moderately differentiated. The second type is *diffuse adenocarcinoma*, which has small groups of tumour cells without formation of glandular structure. Signet ring cell carcinoma has mucin-containing cells with nuclei at the periphery. It has extensive newly formed stroma making identification of cells difficult in standard haematoxylin and eosin sections.

It is important to realize that gastric carcinoma is a biologically heterogeneous disease involving numerous genetic and epigenetic alterations. A very small proportion of gastric cancers can be caused by a specific germline mutation of the *CDH1* gene. This cancer is developed through mismatched processes that begin with *H. pylori*-induced atrophic gastritis. Epstein–Barr virus is another infectious cause of gastric cancer. Mutations of tumour protein B53 -catenin genes occur early in the development of gastric cancer and contribute to gastric carcinogenesis. Also, a significant number of gastric cancers show loss of Runx3 due to homozygous deletion and hypermethylation of the promoter region. Cdx2 has been shown in precancerous lesions as well as gastric cancer. Gastric cancer with microsatellite instability is also a well-defined subset exhibiting distinct cardinal pathological features.

Diagnosis

Clinical assessment

There are no typical signs suggestive of gastric cancer. Patients can present with a variety of symptoms including epigastric pain, loss of appetite, weight loss, bleeding, obstruction or tiredness due to anaemia and perforation. It is important to realize that the majority of the symptoms are in advanced disease. Approximately 70% of patients with early gastric cancer have symptoms of uncomplicated dyspepsia with no associated anaemia, dysphagia or weight loss. The use of alarm symptoms to select patients for endoscopy may cause patients with localized disease to be overlooked. The recommendation in the UK is that patients with dyspepsia who are older than 55 years of age with persistent new symptoms or those with alarm features at any age should undergo endoscopy, as clinical diagnosis alone is inaccurate to distinguish between cancer and non-cancerous gastric disorders. All patients deemed to be at risk with dyspepsia should be considered for endoscopy, even though the overall detection rate is 1–3%.

The clinical examination of a patient with gastric cancer may reveal anaemia, positive Virchow lymph node or abdominal mass. Clinical examination is only positive in advanced disease. Routine blood and biochemistry are important to check for anaemia, dehydration or abnormal liver function.

Endoscopy

Upper gastrointestinal endoscopy with biopsies remains the gold standard for the diagnosis of gastric cancer. The reporting of endoscopic findings should be in a standard manner detailing the description of the lesion, satellite lesions, the distance from either the gastro-oesophageal junction or pyloris and the location in relation to the greater and lesser curvature. Thorough endoscopic examination is essential. Upper gastrointestinal tract endoscopy should not be regarded as an easy procedure performed by non-experienced endoscopists. Some studies found that about 10% of gastric cancers were missed at an earlier endoscopy in Western countries. It is important to realize that PPIs result in symptomatic improvement in patients with gastric cancer and also result in a degree of healing of ulcers due to gastric cancer.

Staging of gastric cancer

CT scan

The initial staging of gastric cancer should include CT scan with multiple planar reconstruction of the thorax, abdomen and pelvis to determine the presence of metastatic disease, the extent of lymph node involvement and local tumour invasion. The scan should be done with intravenous contrast using a protocol to distend the stomach. One litre of water can be used and gas-forming granules can be administered prior to scanning to achieve maximum distension in order to facilitate the delineation of gastric lesions. Studies have reported that CT scan has sensitivity between 92% and 95% and a specificity of 80–100% for T staging and a sensitivity of 62–90% and specificity of 50–87% for N status.

Endoscopic ultrasound

The main indications for EUS in gastric cancer are to (1) determine the extent of the proximal gastric lesion into the gastro-oesophageal junction; (2) confirm early lesions as these may need endoscopic submucosal resection or proceeding to surgery without neoadjuvant chemotherapy; and (3) examine invasion to the pancreas and other structures in selected cases. The reported sensitivity of EUS for T staging of gastric cancer is between 77% and 100% with a specificity of 67–95%. The reported sensitivity for N staging is between 30% and 96% with specificity of 48–98%. Nevertheless, EUS is not considered a routine investigation for all gastric cancer.

Staging laparoscopy

Laparoscopy with contact ultrasonography provides direct visualization of low-volume peritoneal and liver disease as well as assessing the extent of the tumour in the stomach. Peritoneal lavage and cytology is an integral component of laparoscopy to ensure the absence of small-volume peritoneal disease. The presence of positive peritoneal cytology will change the intent of management into palliative and is associated with poor long-term prognosis. Laparoscopy will upstage the disease in about 5–10% of patients because of peritoneal and liver lesions not seen on CT scan.

PET scan

PET scan if available may upstage patients with gastric cancer but it can be negative, especially in patients with mucinous and diffuse tumours. A PET scan may show distant metastatic disease, especially in patients with equivocal CT scan findings. A PET scan is not a routine investigation in staging gastric cancer.

Summary of preoperative investigations:

- routine bloods and biochemistry investigations such as full blood count, urea and electrolytes, and liver function test
- upper gastrointestinal tract endoscopy and biopsy to confirm the nature of gastric cancer
- staging of the disease with CT scan and staging laparoscopy
- EUS and PET scan may be done in selected cases.

Lymph node stations in gastric cancer

Description of lymph node (LN) stations (Japanese Gastric Cancer Association 2011) (Figure 23.25).

1	Right paracardial LNs, including those along the first branch of the ascending limb of the left gastric artery
2	Left paracardial LNs including those along the oesophagocardiac branch of the left subphrenic artery
3a	Lesser curvature LNs along the branches of the left gastric artery
3b	Lesser curvature LNs along the 2nd branch and distal part of the right gastric artery
4sa	Left greater curvature LNs along the short gastric arteries (perigastric area)
4sb	Left greater curvature LNs along the left gastroepiploic artery (perigastric area)
4d	Right greater curvature LNs along the 2nd branch and distal part of the right gastroepiploic artery
5	Suprapyloric LNs along the 1st branch and proximal part of the right gastric artery
6	Infrapyloric LNs along the first branch and proximal part of the right gastroepiploic artery down to the confluence of the right gastroepiploic vein and the anterior superior pancreatoduodenal vein
7	LNs along the trunk of the left gastric artery between its root and the origin of its ascending branch
8a	Anterosuperior LNs along the common hepatic artery
8p	Posterior LNs along the common hepatic artery
9	Coeliac artery LNs
10	Splenic hilar LNs including those adjacent to the splenic artery distal to the pancreatic tail, and those on the roots of the short gastric arteries and those along the left gastroepiploic artery proximal to its 1st gastric branch
11p	Proximal splenic artery LNs from its origin to halfway between its origin and the pancreatic tail end
11d	Distal splenic artery LNs from halfway between its origin and the pancreatic tail end to the end of the pancreatic tail
12a	Hepatoduodenal ligament LNs along the proper hepatic artery, in the caudal half between the confluence of the right and left hepatic ducts and the upper border of the pancreas
12b	Hepatoduodenal ligament LNs along the bile duct, in the caudal half between the confluence of the right and left hepatic ducts and the upper border of the pancreas
12p	Hepatoduodenal ligament LNs along the portal vein in the caudal half between the confluence of the right and left hepatic ducts and the upper border of the pancreas
13	LNs on the posterior surface of the pancreatic head cranial to the duodenal papilla
14v	LNs along the superior mesenteric vein
15	LNs along the middle colic vessels
16a1	Para-aortic LNs in the diaphragmatic aortic hiatus
16a2	Para-aortic LNs between the upper margin of the origin of the coeliac artery and the lower border of the left renal vein
16b1	Para-aortic LNs between the lower border of the left renal vein and the upper border of the origin of the inferior mesenteric artery
16b2	Para-aortic LNs between the upper border of the origin of the inferior mesenteric artery and the aortic bifurcation
17	LNs on the anterior surface of the pancreatic head beneath the pancreatic sheath
18	LNs along the inferior border of the pancreatic body
19	Infradiaphragmatic LNs predominantly along the subphrenic artery

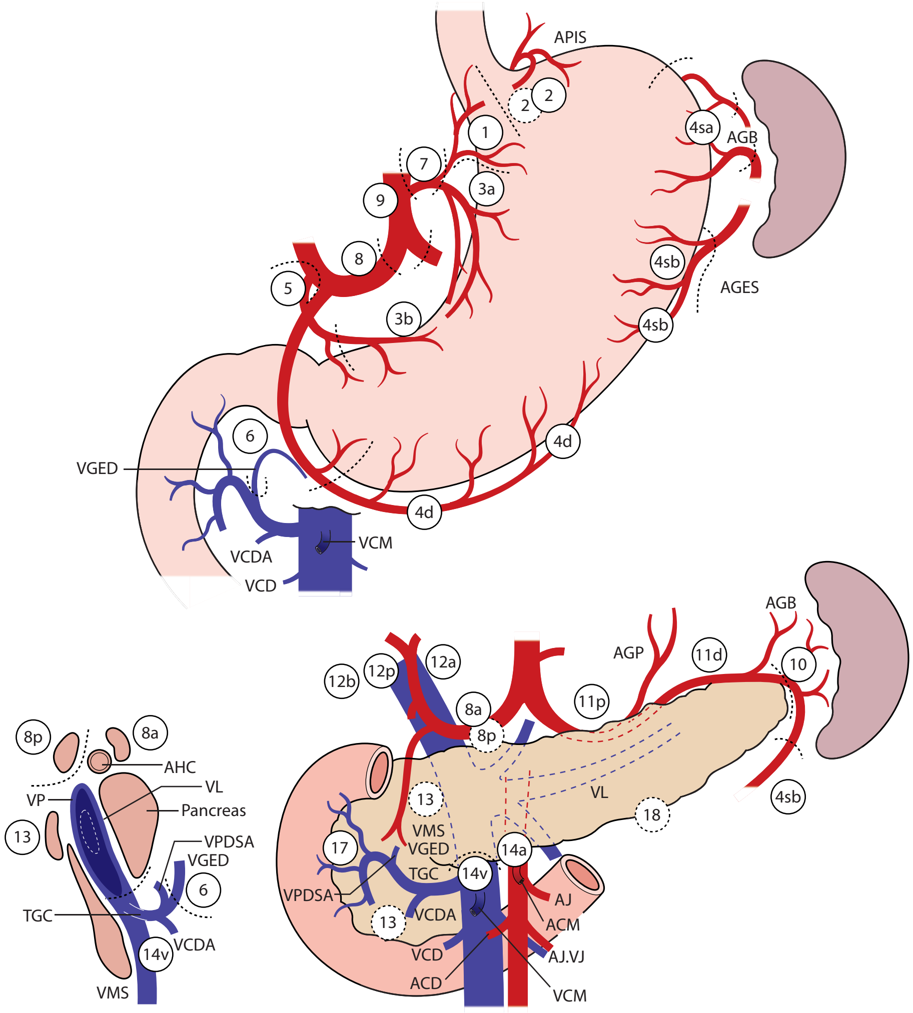


Figure 23.25 Description of lymph node stations (Japanese Gastric Cancer Association 2011). Please refer to text for key. AGB, ACM, AGP, AHC, AJ, TGC, VCD, VCDA, VCM, VGED, VMS, VPDSA.

20	Paraoesophageal LNs in the diaphragmatic oesophageal hiatus
110	Paraoesophageal LNs in the lower thorax
111	Supradiaphragmatic LNs separate from the oesophagus
112	Posterior mediastinal LNs separate from the oesophagus and the oesophageal hiatus

Management of gastric cancer (Figure 23.6)

Multidisciplinary team

Gastric cancer should be treated by a multidisciplinary team. The core members of the team include surgeons, gastroenterologists, oncologists, pathologists, interventional radiologists, palliative team and specialist nurses. All cases of gastric cancer should be discussed by the team to ensure adequate staging and an appropriate management plan. It has been shown that the workload of the surgeon is important for the outcome of gastric cancer in terms of postoperative morbidity and mortality because high-volume workload drives the processes of care and consequently improves early outcomes. The effect of surgeons' case loads on long-term outcome also depends on the quality of the surgery performed by surgeons. Long-term survival is a direct function of the stage of the disease and the extent of lymphadenectomy as a marker for surgical quality. Nevertheless, case workload and the extent of lymphadenectomy vary widely in clinical practice. Hospital case load is as important as surgeons' case load, although best results are seen in hospitals in which many patients are treated by surgeons with much experience.

Endoscopic treatment for early gastric cancer

Endoscopic resection has been accepted as a curative modality for early gastric cancer. Interventional mucosal resection was introduced in 1994 initially for early flat-type gastric cancer. The flat lesion is changed to a polypoid lesion by submucosal injection of a mixture of normal saline with diluted epinephrine (adrenaline) and dissected by grasper forceps and electrosurgical snare through a two-channel endoscope. This method has the advantage of easy technique, short procedure time and fewer complications, with a complete resection rate around 50–70%.

In 1988 the endoscopic mucosal resection (EMR) precutting method was developed in which, after hypertonic saline was diluted, epinephrine (adrenaline) is injected to the submucosal

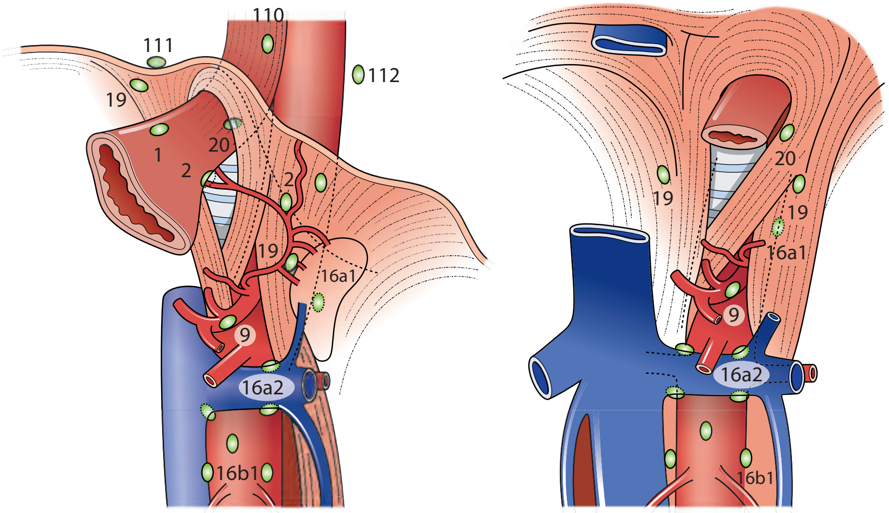


Figure 23.26 Diagrammatic representation of gastric station lymph nodes.

layer around the lesion; circumferential incision around the lesion with a needle knife and final resection of the lesion are performed with electrosurgical snares. This method has enabled complete envelope resection of the region rather than conventional EMR.

In 1992 EMR by a cap-fitted method was developed in which *en bloc* resection was performed with electrosurgical snare after suction of the lesion by a cap-fitted endoscope. Although this method was easier and safer than conventional EMR for *en bloc* resection of small early gastric cancer, *en bloc* complete resection with this technique was very difficult for larger sized lesions.

Endoscopic submucosal dissection (ESD) has been developed to overcome the limitations of EMR. In endoscopic submucosal resection circumferential incision and direct dissection of the submucosal layer of early gastric cancer has been possible with various knives. This enables complete resection, even in large or the depressed type of early gastric cancer. Various knives have been introduced for safe and complete ESD for early gastric cancer.

The rationale behind EMR is to maintain the quality of life without compromising long-term survival. This means that endoscopic resection of early gastric cancers should only take place in cases without lymph nodes or distant metastases to ensure that long-term survival rates are not inferior to surgical resection. Conventional indications for endoscopic resections are (1) differentiated adenocarcinoma confined to mucosa; (2) elevated lesion less than 2 cm; and (3) depressed type without an ulcer in less than 1 cm. With these indications the tumour rarely metastasizes to regional lymph nodes or distant organs. In the area of endoscopic submucosal resection, most limiting factors have been overcome in complete resection and the indications for ESD are decided by the risk of lymph node or distant metastases.

Early gastric cancer is defined as tumour of the stomach confined to the mucosa or submucosa. For this tumour, the risk of lymph node metastasis is very low. Systematic review of the effectiveness and safety of endoscopic resection was identified in a non-randomized trial comparing endoscopic resection treatment. The results of studies of endoscopically treated patients showed disease-specific survival at 5 years and 10 years of more than 95%. Incidence of local recurrence is only 6%, and the chance of complication compares favourably with surgery (0.6% perforation and 14% bleeding). Prophylactic eradication of *H. pylori* after EMR significantly reduces the development of metachronous tumours. Risk of lymph node metastasis rises with increased tumour stage. For early gastric cancer the risk of lymph node metastasis is between 2% and 5% in patients with mucosal cancer and 10–20% for those with submucosal cancer.

The following management of early gastric cancer by endoscopic resection is according to the Japanese Gastric Cancer Guidelines in 2010. In EMR, the lesion together with the surrounding mucosa is lifted by submucosal injection of saline and removed using high-frequency electrosurgery steel snare. In ESD, the mucosa surrounding the lesion is circumferentially incised using a high-frequency electric knife and the submucosal layer is dissected from the proper muscle layer. The lesion is histologically classified as either differentiated or undifferentiated type. The former includes capillary adenocarcinoma and tubular adenocarcinoma. The latter includes poorly differentiated adenocarcinoma, signet ring cell carcinoma and mucinous adenocarcinoma. The tumour consisting of both differentiated- and undifferentiated-type carcinoma is classified according to the quantitative predominance. Diagnosis of ulcerative findings is principally based on histological evidence of ulcerative findings.

The standard indication for EMR or ESD is differentiated-type adenocarcinoma without ulcerative findings of which the depth of invasion is clinically diagnosed as T1a and the diameter is less than 2 cm. Curative resection is judged when all the following conditions are fulfilled: *en bloc* resection, tumour size less than 2 cm, histologically or differentiated type, pT1a, negative horizontal margin, negative vertical margin and no lymph or vascular infiltration. There are expanded criteria for ESD which is regarded by the Japanese Cancer Association as investigational at this stage. After endoscopic curative resection, *H. pylori* should be tested and if positive eradicated. Follow-up by annual endoscopy is recommended.

Surgery

Surgery is the mainstay for curative treatment for gastric cancer.

Curative surgery

1. Standard gastrectomy is the principal surgical procedure performed with curative intent. It involves resection of at least two-thirds of the stomach with D2 lymph node dissection.
2. In a non-standard gastrectomy the extent of gastric resection and/or lymphadenectomy is altered according to the tumour characteristics.
   - In modified surgery the extent of gastric resection and/or lymphadenectomy is reduced compared to standard surgery.
   - In extended surgery, gastrectomy with combined resection of adjacent involved margin or gastrectomy extended lymphadenectomy beyond D2 resection.

Non-curative surgery

Palliative surgery is indicated for bleeding or obstruction in patients with advanced gastric cancer with unresectable metastases. Palliative surgery usually aims to relieve the symptoms. Palliative gastrectomy or gastrojejunostomy is selected depending on tumour characteristics and surgical risk. However, palliative surgery is an extremely uncommon procedure in Western centres because of the advances of pyloric stents and oncological management.

Extent of gastric resection

A sufficient gastric resection margin should be ensured when determining the resection line in operations with curative intent. A proximal margin of at least 3 cm is recommended for T2 or deeper tumours with expansive growth pattern types I and II, and 5 cm is recommended for those with expansive growth pattern types III and IV. When tumours invade the gastro-oesophageal junction a 5 cm margin may not be achieved, and therefore frozen section is recommended to ensure R0 resection. The standard surgical procedure for clinically node positive or T2 to T4a tumours is either total or distal gastrectomy. Distal gastrectomy is selected when a satisfactory proximal margin can be obtained. Pancreatic invasion by tumour requiring pancreaticosplenectomy necessitates total gastrectomy regarding the tumour location. Total gastrectomy with splenectomy should be considered for tumours that are located along the greater curvature and harbour metastasis to 4sb lymph nodes even if the primary tumour could be removed by distal gastrectomy.

For T1N0 tumours, gastric resection can be modified as pyloric preserving gastrectomy for tumours in the middle portion of the stomach with distal tumour at least 4 cm proximal to the pylorus and proximal gastrectomy for proximal tumours more than half of the distal stomach being preserved.

For a proximal gastric tumour invading the gastro-oesophageal junction, gastrectomy with lower oesophagectomy is recommended. If the tumour invading the oesophagus is less than 2 cm, a total abdominal approach with transhiatal resection of the lower oesophagus and lower mediastinal lymph nodes can be achieved. If the tumour extends more than 2 cm into the oesophagus, patients usually require left thoracolaparotomy. Extensive gastric tumour involving the junction more than 2 cm usually requires total gastrectomy with lower oesophagectomy with Roux-en-Y oesophagojejunal reconstruction.

Omentectomy

Removal of gastric omentum is usually integrated in standard gastrectomy for T3 or debulked tumours. Bursectomy has a survival benefit in T3 and T4a tumours. For tumour penetrating the serosa or the posterior gastric wall, complete bursectomy including the pancreatic capsule is performed with the aim of removing microscopic tumour deposits in the lesser sac. Bursectomy should be avoided in T1–T2 tumours to minimize chances of injury to the pancreas.

Combined resection

It is essential to obtain R0 resection and therefore, for tumour invading adjacent organs, combined resection may be necessary.

Laparoscopic surgery

The benefits of laparoscopic surgery in benign disease and colorectal cancer surgery have encouraged surgeons to adopt this in gastric cancer. There is a wide variation in the indications of laparoscopy between the Western countries and Japan. In Japan, although the laparoscopic approach is widely performed for early distal gastric cancer, it is still considered an investigational procedure. The long-term oncological value of laparoscopic distal gastrectomy is not proven in the randomized control trial and we are still awaiting the outcomes in those trials in Japan and Korea. Small randomized control trials showed the potential benefit in terms of general postoperative outcomes, although trials have shown no difference in major postoperative complications. Some trials documented shorter hospital stay with the laparoscopic approach. It is important to recognize that the aim of surgery for gastric cancer is long-term cure. If the surgeon feels that the laparoscopic approach may compromise the oncological quality of the resection, surgeons should revert to the traditional open approach.

Lymph node dissection

The extent of systemic lymphadenectomy is defined according to the type of gastrectomy indicated.

Total gastrectomy

- D1: lymph node stations 1–7
- D1+: D1 + lymph node stations 8a, 9, 11p
- D0: lymphadenectomy less than D1
- D2: D1+ 8a, 9, 10, 11p, 11d, 12a

For tumours invading the oesophagus, D1+ includes lymph node station 110 (lower thoracic, paraoesophageal nodes), and D2 includes lymph node stations 19, 20, 110 and 111.

Distal gastrectomy

- D1: lymph node stations 1, 3, 4sb, 4d, 5, 6, 7
- D1+: D1 + lymph node stations 8a and 9
- D0: lymphadenectomy less than D1
- D2: D1 + lymph node stations 8a, 9, 11p, 12a

Pyloric preserving gastrectomy

- D1: lymph node stations 1, 3, 4sb, 4d, 6, 7
- D1+: D1 plus lymph node stations 8a and 9
- D0: lymphadenectomy less than D1

Proximal gastrectomy

- D1: lymph node stations 1, 2, 3a, 4sa, 4sb and 7
- D1+: D1 + lymph node station 8a, 9, 11p
- D0: lymphadenectomy less than D1

In principle D1 or D1+ lymphadenectomy is indicated for T1N0 tumours and D2 is indicated for N+ or T2–T4 tumours. Because the pre- and intraoperative diagnosis of lymph node metastasis may need revision, D2 lymphadenectomy should be performed whenever nodal involvement is suspected. D1 lymphadenectomy is indicated for a T1a tumour that does not meet the criteria for EMR and for T1bN0 tumours that are not histologically of the differentiated type and 1.5 cm or smaller in diameter. A D1+ lymphadenectomy is indicated for T1N0 tumours other than the above. D2 lymphadenectomy is indicated for potentially curable T2–T4 tumours as well as T1N+ tumours.

D2+ lymphadenectomy: gastrectomy with extended lymphadenectomy beyond D2 is classified as a non-standard gastrectomy. The benefit of prophylactic para-aortic lymphadenectomy was not proven in the Japanese randomized control trial.

Rationale for lymph node dissection

Evidence for D1 dissection

Two randomized trials in the Western countries (MRC trial and the Dutch trial) showed no significant survival benefit of D2 lymphadenectomy. The 5 year survival in the MRC trial was 35% for D1 vs 33% for D2 while in the Dutch trial it was 45% for D1 vs 47% for D2. This equality of long-term survival was associated with higher morbidity and mortality for D2 lymph node dissection. The postoperative mortality was 6.5% vs 13% and 4% vs 10% for D1 vs D2 in the MRC and Dutch trials respectively. The morbidity was 28% vs 46% and 25% vs 43% for the D1 vs D2 gastrectomy in the MRC and Dutch trials respectively. The conclusion from both trials favoured D1 gastrectomy because of similar long-term survival while D1 has a better early postoperative outcome. Therefore the value of D2 dissection was not proven by these trials. The trials also showed that D2 was technically a more demanding operation. Nevertheless, the results of those trials were inferior compared with specialized Western centres.

There are several factors responsible for the outcomes of these trials. The annual average number of cases per hospital was 1 and 1.5 resections for the Dutch and MRC trials respectively. The surgical inexperience and the learning curve occurring throughout the trials may have impacted on the results of both arms of the trial. In the Dutch trial, the contamination rate of 52% in D1 resection (performance of more extensive dissection than specified) and a non-compliance rate of 84% in D2 gastrectomy (performance of less dissection than specified) led to partial homogenization of the groups undermining the likelihood of detecting any potential advantage to D2 dissection. Similarly in the MRC trial only 75% of D2 group had an adequate lymphadenectomy. Also, the high 30 day mortality offsets the effects of D2 dissection in the long term. This inadequate surgical performance was coupled with inexperience in managing postoperative complications. For instance, the mortality rate was 41.3% and 20% for patients who developed anastomotic leakage and pancreatic fistula, respectively, in the Dutch trial.

The evidence drawn from the randomized control trials conducted more than 20 years ago may not be applicable to modern surgical treatment for gastric cancer, as practice has changed significantly. Nowadays, surgery for invasive gastric cancer is carried out in regional high-volume centres by multidisciplinary teams with careful case selection and appropriate high-quality postoperative support. The reported outcomes of D2 cohort studies from several European high-volume centres have shown a lower postoperative mortality and a better 5 year survival than those reported in the MRC and Dutch trials. The principal investigator in the MRC D1 vs D2 gastrectomy randomized control trial believes that the results of this trial are no longer a sustainable argument against D2 gastrectomy in modern surgery for invasive gastric cancer. The fact is that every randomized control trial has a limited lifespan during which its data should influence surgical practice and, although we should adhere firmly to the practice of evidence-based medicine, we should also recognize that like everything else in life including results of randomized control trials have a sell-by date and that pertaining to the MRC D2 gastrectomy has now expired.

Evidence for D2 lymphadenectomy

- Biological behaviour of cancer spread
- Pattern of recurrence
- Value of local control
- Survival after D2 gastrectomy dissection
- Morbidity and mortality after D2 dissection

The incidence of lymph node metastasis with gastric cancer highlights the biological behaviour of gastric cancer. The data from the National Cancer Centre, Tokyo, showed that the incidence of positive lymph nodes with pT2 is 46–64%, for pT3 is 78.9% and for pT4 is 89.8%. This means that limited lymphadenectomy is likely to leave tumour in the non-resected stations.

With regard to recurrence after limited resection, local recurrence and/or lymph node metastasis are the only failures in 53.7% of the failure group when localized peritoneal failure is included, and as a component of failure in 87.8% of patients. By contrast, lymph node recurrence after D2 lymphadenectomy in the National Cancer Centre, Tokyo, is 7%. This contrast highlights the value of lymphadenectomy in minimizing local disease recurrence. It is known that local recurrence is very difficult to palliate and compromises the quality of life of patients. Good local control is essential in the management of cancers and R0 is the most important prognostic factor after gastrectomy, as shown in many clinical studies. The postoperative chemotherapy randomized control gastric trial (IT-116/SWOG) showed that chemoradiotherapy improves survival after limited gastrectomy. It also showed that surgical undertreatment was an independent prognostic factor in retrospective analysis of the trial. Subgroup analysis showed that chemoradiotherapy does not improve survival following D2 gastrectomy. These data showed that D0/D1 are inadequate for local control requiring chemoradiotherapy to improve the outcome.

The value of D2 gastrectomy has been proven in several D2 cohort studies from European centres which showed postoperative mortality less than 5% and 5 year survival of 50–60%, approaching the Japanese standard. This encouraged surgeons to adapt extended lymphadenectomy with improved results. Also the 15 year follow up of the Dutch trial showed a significant survival benefit of extended lymphadenectomy. The benefit of D2 gastrectomy was proved in the Taiwan randomized control trial, which showed overall long-term survival benefit with extended lymphadenectomy compared with D1 dissection. No postoperative death occurred in either group. However, D2 gastrectomy caused increased morbidity. It is important to note that distal pancreatectomy and splenectomy should not be routinely done in D2 gastrectomy because they increase the morbidity and mortality of the operation.

Perioperative chemotherapy

The MAGIC trial demonstrated that a treatment plan of three cycles of preoperative and postoperative epirubicin (E) 50 mg/m^2, cisplatin (C) 60 mg/m^2 and continuous intravenous infusion of 5-fluorouracil (F) 200 mg/m^2 per day (ECF) significantly improved 5 year survival from 23% with surgery alone to 36.3%. The main non-haematological toxicity was alopecia, nausea and vomiting. The results were supported by the FFCD trial. This perioperative approach has been adopted as the standard of care in most of the UK and parts of Europe. Because of the non-inferiority of capecitabine (X) compared with 5FU in advanced gastric cancer and because it obviates the need for an indwelling central venous access device, many centres use ECX in the perioperative setting.

Postoperative chemoradiotherapy

A North American intergroup randomized trial demonstrated that five cycles of postoperative chemotherapy with 5FU/leucovorin before, during and after radiotherapy (45 Gy in 25 fractions over 5 weeks) resulted in about a 15% improvement in overall survival. Although this treatment is considered to be the standard in the USA, it has not gained acceptance in Europe because of the late toxicity with abdominal chemoradiotherapy and because of the quality of surgery used in this trial. Only 10% of trial participants underwent D2 dissection, whereas the beneficial effect of postoperative chemoradiotherapy appeared greatest in patients who received D1 dissection (36%) or less than D1 (54%) resections. Although this study suggested that postoperative chemoradiation compensates for suboptimum surgery, a large non-randomized observational study suggested a potential benefit for postoperative chemoradiation after D2 dissection. The results of the randomized study from Korea in which patients were given either capecitabine and cisplatin chemotherapy or capecitabine and cisplatin chemoradiotherapy after D2 resection are awaited with great interest. In a meta-analysis, postoperative chemoradiotherapy was reported to improve survival significantly. Modern high-precision radiation techniques and more intensified chemoradiation regimes are more likely to further improve the results of postoperative chemoradiation.

Adjuvant chemotherapy

Several meta-analyses have suggested a small survival benefit for adjuvant chemotherapy. However, there is considerable variation between the treatment regimes and outcomes between Western and Asian populations. The Japanese adjuvant chemotherapy trial of Ts1 for gastric cancer demonstrated a significant benefit in overall survival for patients receiving 12 months of S1 (an oral fluoroamidine) monotherapy compared with observation after curative D2 gastrectomy for node-positive disease. No similar trial has been carried out in Western populations. Adjuvant chemotherapy alone is not a standard practice in the UK, although it may be used in patients with high risk of recurrence who did not have neoadjuvant chemotherapy.

Neoadjuvant chemoradiotherapy

Neoadjuvant chemoradiation with cisplatin and etoposide after induction chemotherapy with cisplatin, 5FU and leucovorin is feasible for locally advanced carcinoma at the gastro-oesophageal junction. However, the study was underpowered due to poor recruitment. There are small randomized controlled trials which showed that neoadjuvant chemoradiotherapy results in improved overall 3 year and 5 year survival. A meta-analysis comparing surgery with surgery preceded by radiotherapy showed significant improvement in 3 and 5 year survival without a rise in postoperative mortality. Although these studies showed advantages of preoperative radiotherapy and surgery, further studies are unlikely because current research is directed towards either perioperative chemotherapy or postoperative chemoradiotherapy. Theoretically neoadjuvant chemoradiation could be a good strategy because chemotherapy would not be delayed by postoperative recovery; the treatment target area is easy to identify because the tumour and stomach are still in the normal position, with good vascularization and oxygenation of tumour tissue without major anatomical deviation; and tumour downsizing could facilitate surgery. A disadvantage is that the pathological staging is unavailable.

Staging of gastric cancer

Rules for classification

The classification applies to carcinomas. There should be histological confirmation of the disease. A tumour the epicentre of which is within 5 cm of the oesophagogastric junction and also extends into the oesophagus is classified and staged according to the oesophageal scheme. All other tumours with an epicentre in the stomach greater than 5 cm from the oesophagogastric junction, or those within 5 cm of the junction without extension into the oesophagus, are staged using the gastric carcinoma scheme.

The following are the procedures for assessing T, N and M categories:

- T categories: physical examination, imaging, endoscopy and/or surgical exploration
- N categories: physical examination, imaging and/or surgical exploration
- M categories: physical examination, imaging and/or surgical exploration.

Regional lymph nodes

The regional lymph nodes of the stomach are the perigastric nodes along the lesser and greater curvatures, the nodes along the left gastric, common hepatic, splenic and coeliac arteries and the hepatoduodenal nodes. Involvement of other intra-abdominal lymph nodes such as retropancreatic, mesenteric and para-aortic is classified as distant metastasis.

TNM clinical classification (version 7)

TNM classification (version 7) showed a better prognostic predictive ability than TNM (version 6) in both T and N status.

T – Primary tumour

- Tx: Primary tumour cannot be assessed
- T0: No evidence of primary tumour
- Tis: Carcinoma *in situ*: intraepithelial tumour without invasion of the lamina propria, high-grade dysplasia
- T1: Tumour invades lamina propria, muscularis mucosae or submucosa
  - T1a: Tumour invades lamina propria or muscularis mucosae
  - T1b: Tumour invades submucosa
- T2: Tumour invades muscularis propria
- T3: Tumour invades subserosa
- T4: Tumour perforates serosa or invades adjacent structures
  - T4a: Tumour perforates serosa
  - T4b: Tumour invades adjacent structures
- The adjacent structures of the stomach are the spleen, transverse colon, liver, diaphragm, pancreas, abdominal wall, adrenal gland, kidney, small intestine and retroperitoneum.
- Intramural extension to the duodenum or oesophagus is classified by the depth of greatest invasion in any of these sites, including the stomach.
- Tumour that extends into gastrocolic or gastrohepatic ligaments or into greater or lesser omentum, without perforation of visceral peritoneum, is T3.

N – Regional lymph nodes

- Nx: Regional lymph nodes cannot be assessed
- N0: No regional lymph node metastasis
- N1: Metastasis in 1 or 2 regional lymph nodes
- N2: Metastasis in 3–6 regional lymph nodes
- N3: Metastasis in 7 or more regional lymph nodes
  - N3a: Metastasis in 7–15 regional lymph nodes
  - N3b: Metastasis in 16 or more regional lymph nodes

M – Distant metastasis

- M0: No distant metastasis
- M1: Distant metastasis
  - Distant metastasis includes peritoneal seeding, positive peritoneal cytology and omental tumour not part of continuous extension.

In *pTNM Pathological Classification,* the pT and pN categories correspond to the T and N categories.

- pN0: Histological examination of a regional lymphadenectomy specimen will originally include 16 or more lymph nodes. If the lymph nodes are negative, but the number ordinarily examined is not met, classify as pN0.
- pM1: Distant metastasis microscopically confirmed.

G histopathological grading

- Gx: Grade of differentiation cannot be assessed
- G1: Well differentiated
- G2: Moderately differentiated
- G3: Poorly differentiated
- G4: Undifferentiated

Stage grouping

- Stage 0: TisN0M0
- Stage 1A: T1N0M0
- Stage 1B: T2N0M0
  - T1N1M0
- Stage IIA: T3N0M0
  - T2N1M0
  - T1N2M0
- Stage IIB: T4aN0M0
  - T3N1M0
  - T2N2M0
  - T1N3M0
- Stage IIIA: T4aN1M0
  - T3N2M0
  - T2N3M0
- Stage IIIB: T4BN0, N1M0
  - T4aN2M0
  - T3N3M0
- Stage IIIC: T4aN3M0
  - T4bN2, N3M0
- Stage IV: Any T, Any NM1

Summary

- T1: Lamina propria (T1a), submucosa (T1b)
- T2: Muscularis propria
- T3: Subserosa
- T4a: Perforates serosa
- T4b: Adjacent structures
- N1: 1 or 2 nodes
- N2: 3–6 nodes
- N3a: 7–15 nodes
- N3b: 16 or more

Quality of pathological examination

As discussed earlier the value of extended lymphadenectomy by the surgeon is recognized. The TNM staging necessitates a minimum of 16 lymph nodes for staging. The authors of this chapter believe that it is important to retrieve the maximum number of lymph nodes from any given specimen.

This is essential for accurate staging as the N staging of the TNM classification depends on the number of lymph nodes involved with metastasis. Ignoring small lymph nodes can be a major cause of staging error. For instance, if nodes 5 mm or less are ignored, 38% of all metastatic nodes will be missed. Downstaging will occur in 15% and 4% of the cases if nodes of less than 6 and 4 mm respectively are ignored. Accurate staging determines disease prognosis and the need for adjuvant oncological treatment in advanced cases. In addition, the lymph node number is a measure for the extent of lymphadenectomy and operative performance. This is particularly important in oncological trials when added value of oncological therapy is evaluated in a combined arm against surgery alone.

Primary gastric lymphoma

A primary gastric lymphoma is generally non-Hodgkin's lymphoma. They represent 5% of gastric malignancies and show an apparently increasing incidence worldwide. Primary and secondary involvement of the stomach is extremely rare in Hodgkin's disease.

Histological characteristics

Primary extranodal non-Hodgkin's lymphoma was taken into consideration for the first time in 1994 with the REAL classification (Revised European American Lymphoma classification). The WHO classification originates directly from the REAL classification and distinguishes over 30 different forms. As a result non-Hodgkin's gastric lymphoma forms a very heterogeneous group of lymphoproliferative tumours with a wide variety of biological and clinical responses to treatment. There are two main histological types: diffuse large B-cell lymphoma and marginal zone B-cell lymphoma of MALT type. Follicular, mantle cell and peripheral T-cell lymphomas represent about 2% of gastric lymphomas.

Pathogenesis of B-cell MALT gastric lymphoma

This is associated with chronic *H. pylori* infection. MALT lymphoma represent 8% of all B-cell lymphomas and approximately 50% of gastric lymphomas. They originate in the stomach from lymphatic tissue that is in association with the mucous membrane, which usually develops following chronic *H. pylori* infection. *H. pylori* is detected in 60% of gastric lymphomas. The presence of *H. pylori* is significantly more frequent in patients with lymphoma restricted to the mucosa and submucosa (76%) than in those with lymphoma invading beyond the submucosa (48%). It is also more common in patients with maltomas than in those with high-grade tumours. The frequency of *H. pylori* positivity is significantly lower in gastric lymphoma cases than in chronic active gastritis (100%) or peptic ulcer disease (91%). In a large case–control study, there was a statistically significant association between previous *H. pylori* infection and the development of primary gastric lymphomas with either low-grade (MALT type) or high-grade diffuse large cells (histology). The correlation between *H. pylori* and lymphomas has been confirmed on biological and molecular levels both through gastric biopsies showing the presence of B lymphocytic clones that will later give origin to the tumour and through the ability to induce growth of the lymphomatous tissue *in vitro* when stimulated by *H. pylori* in culture. Importantly, it is known that *H. pylori* eradication with antibiotics is able to induce remission of gastric MALT lymphomas in the majority of cases.

In MALT lymphomas at least three recurrent chromosomal translations have been identified and together are present in approximately 65% of cases: the t(11:18)(q21) translocation, which causes the formation of the chimaeric fusion chain AP12-MALT1; the t(1:14)(p22:q32) translocation, which causes deregulation of PCL10; and the t(14:18)(q32:q21) translocation, causing the deregulation of MALT1.

Pathogenesis of diffuse large B-cell gastric lymphomas

In a high percentage of extranodal high-grade lymphomas, the oncogene BCL6 (located in chromosome 3q27) is altered due to translocations, somatic hypermutation or deregulation mutations involving the promoter region. These three arrangements induce an overexpression of the gene which seems to predict a better prognosis.

Grade of lymphomas

The grading of gastric lymphoma is important in both prognosis and treatment. Most gastric lymphomas are of B-cell origin and divided into low-grade and high-grade tumours. Low-grade MALT lymphomas are fairly indolent tumours that often remain localized for extended periods of time. High-grade MALT lymphomas perforate and disseminate more rapidly. The frequency of lymph node involvement also correlates with the grade of lymphoma.

Clinical picture

The initial clinical picture is often non-specific and indicative of a gastric or ulcerative process rather than a neoplasm. The most common symptoms are epigastric pain and dyspepsia with nausea and vomiting, anorexia and weight loss. Gastrointestinal haemorrhage occurs in the outset in 20% of patients, whereas gastric occlusion and perforation are quite uncommon. On examination, it is important to palpate all lymph node regions as well as the abdomen for hepatosplenomegaly or organomegaly.

Investigations

Routine laboratory investigations include full blood count, biochemistry, liver function tests and serum protein electrophoresis. Diagnosis is usually established with upper gastrointestinal tract endoscopy. The three main patterns are ulcerative, diffuse infiltration or polypoid mass. Therefore the endoscopic findings can range from gastritis-like and superficial ulcers to diffuse thickening and granularity of mucosal folds and submucosal mass-like effect. CT scan of the chest, abdomen and pelvis is required. Abdominal CT scan abnormalities are found in 70% of patients and include gastric wall thickening and lymphadenopathy. These abnormalities are more common with high-grade lymphoma than low-grade lymphoma

(100% vs 51%). PET is important in staging lymphomas (Box 23.6) as it will demonstrate extragastric metabolically active lymphadenopathy.

Management

Treatment of early-stage H. pylori-positive MALT-type gastric lymphoma

The most widely accepted initial treatment for localized disease is aimed at eradication of *H. pylori* infection using a combination of antibiotics and PPIs. Numerous reports confirmed the effectiveness of antibiotics in sustaining long-term remission in 60–100% of patients with localized *H. pylori*-positive maltoma. There is no evidence to suggest additional treatment with chemotherapy in patients who respond to eradication therapy.

Treatment of patients with advanced or H. pylori-negative MALT-type gastric lymphomas

In general *H. pylori*-negative gastric MALT lymphomas do not regress when treated with antibiotics, but some response has been reported in early-stage cases and therefore due to the indolent nature of the disease and taking into account more aggressive approaches used in antibiotic-refractory patients, an initial therapeutic attempt might be considered for *H. pylori*-negative patients with MALT lymphomas confined to the stomach. Excellent results in local disease control have been reported using radiotherapy on the stomach and perigastric lymph nodes.

Up to 90% of patients receiving radiotherapy alone achieved a complete response with 5 year disease-free survival and an overall survival rate of 98% and 77% respectively. There are now sufficient data to demonstrate that conservative therapy permits a better quality of life with no negative impact on survival rates. Therefore, surgery no longer has a central role in the therapeutic strategy and is reserved only for carefully selected cases in which alternative treatments are not possible. Surgery is also used for the treatment of complications.

In all patients with disease that has spread, systemic therapy similar to that of advanced lymphomas must be taken into consideration. Treatment options include chemotherapy and the use of monoclonal antibodies.

BOX 23.6 International Workshop Staging System for gastrointestinal non-Hodgkin's lymphoma

Stage I Tumour confined to the gastrointestinal tract
- Single primary site or multiple non-continuous lesions

Stage II Tumour extending to the abdomen for the primary gastrointestinal site
- Nodal involvement:
- II1 local (paragastric or paraintestinal)
- II2 distant (mesenteric para-aortic, paracaval, pelvic, inguinal)

Stage III Penetration of the serosa to involve adjacent organs or tissue

Stage IV Disseminated extranodal involvement or gastrointestinal tract lesion with supradiaphragmatic nodal involvement

Treatment of high-grade gastric lymphoma

Diffuse large -cell non-Hodgkin's lymphoma represents the most common histological type of lymphoma. The treatment is based on aggressive polychemotherapy that is usually combined with rituximab. The need for surgery for these cases has completely disappeared because of the advances in chemotherapy and staging modality. There is no need for surgery as a staging procedure. The same guidelines followed for aggressive lymphomas can be applied to gastric lymphoma with aggressive histology.

Follow-up

Regular follow-up is mandatory as negative biopsy cannot exclude the presence of aggressive diffuse large -cell lymphoma and should require proactive therapy with curative intent. EUS with multiple biopsies should be done 3–6 months following *H. pylori* eradication to evaluate lymphoma regression, repeated every 6 months for 2 years. For aggressive histology, CT scan every 3 months for 2 years is recommended. When performed and it shows a positive diagnosis, the PET scan should be repeated.

In summary, most strategies for the treatment of gastric lymphoma no longer include surgical resection in primary treatment and surgery is reserved for the management of complications or unique cases or locally persistent disease.

Gastric gastrointestinal stromal tumour

Epidemiology

GIST is a rare sarcomatous tumour of the gastrointestinal tract with an incidence of 10–20/million per annum. It occurs most frequently in the submucosal connective tissue of the stomach (50–70%). Additionally, the small intestine (20–30%) and, more rarely, oesophagus, colon, rectum, omentum and mesentery are also sites of GIST formation. There is a higher propensity in men with a median age of diagnosis between the fifth and seventh decades.

GISTs are the result of mutations of either the *c-kit* proto-oncogene (encoding the cytokine receptor protein CD117/KIT protein) or *PDGFRA* gene (which encodes the cell surface tyrosine kinase receptor, -type platelet-derived growth factor receptor). Eighty per cent of mutations occur in the *c-kit* gene whereas 5–10% occur in the *PDGFRA* gene. Ten to 15% of cases of GISTs have no detectable gene mutation; however, the diagnosis is not excluded by their absence.

Pathology

Histologically, GISTs are predominantly composed of spindle-shaped tumour cells (70%) but some can show evidence of epithelioid tumour cells (20%) or even a combination of the two cell types, the so-called 'mixed type' (10%). As a result, standard haematoxylin and eosin staining can make it difficult to differentiate GISTs from other spindle cell tumours such as leiomyoma or leiomyosarcoma as well as carcinomatous and carcinoid lesions especially in the presence of epithelioid-type GISTs.

As a result, although a large part of tissue diagnosis is based on standard histological findings, immunostaining is a necessary investigative component for histological confirmation of a GIST. The KIT/CD117 protein is the most important of these markers. GISTs share a common precursor with the interstitial cells of Cajal, which regulate autonomic gut peristalsis and are found between the longitudinal and circular layer of the muscularis propria. The KIT protein is expressed on these cells and is known to be positive in upwards of 95% of GISTs, but is rarely expressed in other abdominal tumours (but may be positive in non-abdominal tumours such as melanoma, seminoma and breast cancer). Additionally, CD34 (positive in 70%), smooth muscle actin (positive in 40%), PS-100 (positive in 5%) and desmin (positive in 2%) are further considered to be useful immunohistochemical markers when diagnosing GIST.

In cases of KIT-negative GISTs (5%), molecular analysis of *c-kit* and *PDGFRA* is considered a useful adjunct for confirming the diagnosis, with their detection having the potential to predict patient response to the tyrosine kinase inhibitors (TKIs). However this technique is not yet widely available and remains predominantly a research tool.

Clinical picture

Presenting features of gastric GIST are often non-specific and, in the case of small tumours, will most often be found on a routine gastroscopy or imaging (CT, MRI, etc.) of the abdominal cavity in asymptomatic patients for an unrelated pathology. Symptomatic patients will typically have larger tumours. GISTs are known to be soft and friable, and gastric GISTs in particular may present with abdominal pain or acute haemorrhage. As tumours tend to grow outwards into the abdominal cavity without impinging on organs, obstruction is rare. Metastatic spread is via the haematogenous route, most frequently to the liver or peritoneal cavity and patients may occasionally present with signs and symptoms of advanced disease.

Diagnosis

Gastric GIST is most commonly diagnosed via direct endoscopy as a submucosal tumour when undertaking a routine gastroscopy. However, given the small circumferential nature of the lesion and non-specific appearance, it must be distinguished from a wide range of differentials including lipoma, leiomyoma, metastatic tumour, carcinoma and gastric cancer.

EUS and fine-needle aspiration biopsy (FNAB) are relatively novel techniques which are increasingly employed to obtain samples from small tumours (especially those up to 3 cm in size) that may not be amenable to standard endoscopic biopsy. This is a specialized technique which allows visualization of tumour up to its origin in the submucosa and together with echogenicity will provide an accurate assessment of its nature.

CT scanning of the abdominal cavity with intravenous contrast allows tumours to be differentiated from the surrounding structures. CT gives the additional benefit of being widely available and permitting tumour grading and staging (as well as being a modality for follow-up imaging), but the resolution of current scanners for small tumours (<2 cm) is limited.

Classification of gastrointestinal stromal tumour

TNM clinical classification

T – Primary tumour

- Tx: Primary tumour cannot be assessed
- T0: No evidence for primary tumour
- T1: Tumour 2 cm or less
- T2: Tumour more than 2 cm but not more than 5 cm in greatest dimension
- T3: Tumour more than 5 cm but not more than 10 cm in greatest dimension
- T4: Tumour more than 10 cm in greatest dimension

N – Regional lymph nodes

- Nx: Regional lymph nodes cannot be assessed
- N0: No regional lymph node metastasis
- N1: Regional lymph node metastasis

Regional lymph node involvement is rare for GISTs, so that cases in which the nodal status is not assessed clinically or pathologically could be considered N0 instead of Nx or pNx.

M – Distant metastasis

- M0: No distant metastasis
- M1: Distant metastasis

G histopathological grading

Grading for GIST is dependent on the mitotic rate. The mitotic rate of GIST is best expressed as the number of mitoses per 50 high-power fields (HPF) using the 40× objective (total area 5 mm^2 in 50 fields):

- low mitotic rate: 5 or fewer per 50 HPF
- high mitotic rate: over 5 per 50 HPF.

Stage grouping

- Stage 1A: T1, T2N0M0, Low
- Stage 1B: T3N0M0, Low
- Stage II: T1, T2N0M0, High
  - T4N0M0, Low
- Stage IIIA: T3N0M0, High
- Stage IIIB: T4N0M0, High
- Stage IV: Any TN0M0, Any rate
  - Any T: N1M1, Any rate

Summary

- T1: <2 cm
- T2: >2–5 cm
- T3: >5–10 cm
- T4: >10 cm

Surgical management

Surgical resection is the mainstay for the treatment of gastric GIST. Given the relative rarity of GIST, most patients are best managed by multidisciplinary teams in specialist centres, where diagnostic and specialist techniques can be concentrated together with a greater expertise of managing sarcomas.

There are several important features of GISTs which have to be taken into consideration when performing a surgical resection, particularly in the stomach. These are also reflected in the consensus guidance for the surgical management of GIST.

- Primary resection should be complete with margins being macroscopically clear to prevent local or peritoneal recurrence.
- GISTs are soft and friable tumours which can rupture and disseminate. The tumour pseudocapsule should be preserved to prevent spillage. If the procedure is being performed laparoscopically, the tumour should be removed within a bag to prevent intra-abdominal rupture and port site seeding.
- Tumour dissemination is predominantly via the haematogenous route and in particular to the liver and peritoneum. Metastatic spread to local and regional lymph nodes is consequently rare, obviating the need for extensive anatomic and nodal resection as performed in cases of gastric adenocarcinoma. Partial resection to preserve organ function and continuity is the standard treatment.
- GISTs tend to grow into the abdominal cavity, without invading into surrounding tissues and therefore allowing a limited wedge resection to be performed, provided there is no evidence of disseminated disease.
- Although laparotomy has been the traditional approach to resection of gastric GIST, as with all areas of abdominal surgery, the laparoscopic approach is also gaining popularity and becoming the gold standard for tumours up to 4–5 cm diameter. The recognized benefits of laparoscopy are particularly suited to gastric GIST resection, as tumours tend to be excised through a wedge resection, resulting in shorter operative time, reduced blood loss and shorter length of postoperative stay. However, for very large GISTs, surgeons should be very cautious to try the laparoscopic approach as splitting the tumour will result in tumour recurrence.

Gastric GISTs are understood not to metastasize to regional lymph nodes and as a result a limited wedge resection through either an open or laparoscopic approach is preferred, depending on the site of the exact location of the tumour, with the aim of achieving an adequate resection margin and maintaining luminal continuity.

- Around 30% of gastric GISTs are reported to occur on the greater curvature and fundus. This location is amenable to laparoscopic wedge resection. Short-term follow-up data have shown a disease-free survival of more than 90% at 36 months' follow-up in patients who have undergone curative laparoscopic resection, suggesting that the technique is safe.
- GISTs at the gastro-oesophageal junction in particular can be difficult to resect and are associated with a recurrence rate of up to 42%. Owing to the high risk of losing gastro-oesophageal junction patency with a wedge resection, most experts advocate a proximal gastrectomy or even oesophagogastrectomy where there is locally advanced disease via the traditional open approach.
- Small antral tumours may be amenable to wedge resection but larger tumours are typically treated with a distal gastrectomy to maintain gastric continuity.

Medical management

Medical therapy was historically based on chemotherapy which showed little or no response from patients and was associated with an overall 5 year survival rate of only 25%. In the past decade, TKIs have become the mainstay, with ongoing clinical trials showing that tumour growth is controlled in up to 85% of advanced GISTs.

Patients who have non-resectable, metastatic or recurrent disease not amenable to further surgery typically undergo medical management with the TKI imatinib, which blocks the activated proteins produced by the mutated *c-kit* and *PDGFRA* genes. Following the initial dosing of imatinib for 1 month, patients undergo CT or PET scanning to assess response. The medication is continued for as long as possible, as interruption has been associated with a high rate of relapse even in individuals showing complete remission, and is reduced or stopped only if associated with significant side effects (in around 20% of patients) or if responsiveness diminishes over time (this occurs in up to 50% of patients). Intolerant and resistant individuals can be commenced on sunitinib, a second-generation TKI. Patients will generally be scanned every 3 months, to assess response. If a tumour becomes resectable following treatment, then it should be considered for resection. However, at present routine adjuvant and neoadjuvant therapy with TKI is not advocated by any of the GIST consensus groups.

Prognosis

Patients who have had a curative resection are stratified as high, intermediate, low or very low risk as per the guidance of the National Institutes of Health consensus conference (Table 23.5) which links together tumour size and mitotic count under a HPF to give an indication of tissue proliferation potential and therefore the risk of recurrence. Tumour location appears to be important, with gastric GISTs having a more favourable overall outcome than tumours at other sites.

At present there appears to be no consensus as to the frequency of follow-up and reimaging for recurrence, but is generally based on patient risk stratification with CT being the preferred imaging modality as it is quick to perform in regional centres and allows examination of the entire abdominal cavity in a single sitting. The North American National Comprehensive Cancer Network GIST practical guidance recommends CT scanning every 3–6 months for 3 years and yearly thereafter for a further 5 years in high- and intermediate-risk patients. The European Society for Medical Oncology

Table 23.5 Classification according to the National Institutes of Health consensus conference

	Tumour size (cm)	**Mitotic count (per 50 HPF)**
Very low risk	<2	<5/50
Low risk	2–5	<5/50
Intermediate risk	<5	6–10/50
	5–10	<5/50
High risk	>5	5–10/50
	>10	Any mitotic count
	Any size	>10/50

HPF, high-power fields

consensus group advises scanning the same patient group every 3–4 months for the first 3 years and then every 6 months for 5 years. Low-risk individuals are advised to be scanned every 6 months for 5 years. A joint Japanese guideline adopts a middle approach, recommending that patients of high or intermediate risk are followed up with CT scans every 4–6 months, whereas those with a low or very low risk are scanned every 6–12 months.

Overall 5-year survival is highly variable. Patients with primary gastric GISTs who have had a complete resection have a documented survival rate between 20–63% and recurrence rate of 17–76%.

Summary

- GIST is a rare connective tissue tumour occurring predominantly in the stomach (60–70%) and small intestine (10–20%) but may affect any part of the gastrointestinal tract, mesentery and omentum.
- GIST is usually associated with a mutation to the *c-kit* or *PDGFRA* genes encoding receptor tyrosine kinases.
- Presenting symptoms are non-specific but gastric GIST may present with acute haemorrhage given its soft and friable nature.
- Diagnosis is through a multimodality approach using immunostaining and imaging techniques of which EUS with FNAB is particularly useful for small tumours.
- Surgery is the mainstay of treatment although tyrosine kinase inhibitors such as imatinib and sunitinib have an important role in prolonging life in cases of unresectable or metastatic disease.
- Tumour size and mitotic rate are used as a guide to predict the malignant potential of GISTs and divide patients into high-, intermediate-, low- and very low-risk groups.
- Follow-up and regular imaging is recommended approximately every 3–6 months in high- and intermediate-risk cases but varies from region to region.

GUIDE TO FURTHER READING

Allum WH, Blazbey J, Griffin SM, *et al.* Guidelines for the management of oesophageal and gastric cancer. *Gut* 2011;**60**:1449–1472.

Anlauf M, Garbrecht N, Henopp T, *et al.* Sporadic versus hereditary gastrinomas of the duodenum and pancreas: Distinct clinico-pathological and epidemiological features. *World J Gastroenterol* 2006;**14**:5440–5446.

Blay JY, Bonvalot S, Casali P, *et al.* Consensus meeting for the management of gastrointestinal stromal tumors. Report of the GIST consensus conference of 20–21 March 2004, under the auspices of ESMA. *Ann Oncol* 2005; **16**:566–578.

Bonenkamp JJ, Hermans J, Sasako M, *et al.* Extended lymph-node dissection for gastric cancer. *N Engl J Med* 1999;**340**:908–914.

Burak KW, Lee SS, Beck PL. Portal hypertensive gastropathy and gastric antral vascular ectasia (GAVE) syndrome. *Gut* 2001;**49**:866–872.

Casaneuva F, Dieguez C. Ghrelin – a new hormone implicated in the regulation of growth hormone secretion and body energy homeostasis. *Growth Genet Hormones* 2004;**20**:1–7.

Cullen JJ, Kelly KA. Gastric motor physiology and pathophysiology. *Surg Clin North Am* 1993;**73**:1145–1160.

Cunningham D, Allum WH, Stenning SP, *et al.* Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* 2006;**355**:11–20.

Cuschieri A, Fayers P, Fielding J, *et al.* Postoperative morbidity and mortality after D1 and D2 resections for gastric cancer: preliminary results of the MRC randomised controlled surgical trial. The Surgical Cooperative Group. *Lancet* 1996;**347**:995–999.

Demetri GD, Benjamin RS, Blanke CD, *et al.* NCCN Task Force. NCCN Task Force report: management of patients with gastrointestinal stromal tumour (GIST) – update of the NCCN clinical practice guidelines. *J Natl Compr Cancer Net* 2007:**5**;S1–S29.

Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol* 1996;**20**:1161–1181.

Fletcher CD, Berman JJ, Corless C, *et al.* Diagnosis of gastrointestinal stromal tumors: a consensus approach. *Human Pathol* 2002:**33**;459–465.

Gervaz P, Huber O, Morel P. Surgical management of gastrointestinal stromal tumours. *Br J Surg* 2009;**96**:567–578.

Hogben AAM, Schanker LS, Tocco D J, Brodie BB. Absorption of drugs from the stomach II: the human. *J Pharmacol Exp Ther* 1957;**120**:540**–**545.

Japanese Gastric Cancer Association. Japanese Classification of Gastric Carcinoma – 2nd English edition. *Gastric Cancer* 1998;**1**:10–24.

Japanese Gastric Cancer Association. Japanese Gastric Cancer Treatment Guidelines 2010. *Gastric Cancer* 2011;**14**:113–123.

Kwee, RM, Kwee TC. Imaging in local staging of gastric cancer: a systematic review. *J Clin Oncol* 2007:**25**:2107–2116.

Lee WH, Oo JH, Kim CH, *et al.* FDG PET/CT scan in diagnosis of recurrent gastric cancer – Comparison with enhanced CT images**.** *J Nucl Med* 2008;**49**(Suppl. 1):351P.

Macdonald JS, Smalley SR, Benedetti J, *et al.* Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. *N Engl J Med* 2001;**345**:725–730.

McColl KEL, El-Omar E, Gillen D. Interactions between *H. pylori* infection, gastric acid secretion and anti-secretory therapy. *Br Med Bull* 1998;**54**:121–138.

Moglia A, Menciassi A, Dario P, Cuschieri A. Capsule endoscopy: progress update and challenges ahead. *Nat Rev Gastroenterol Hepatol* 2009;**6**:353–362.

National Institute for Health and Clinical Excellence. *Dyspepsia: Management of Dyspepsia in Adults in Primary Care*. Clinical Guideline no.17. London, UK:NICE. Available from: www.nice.org.uk/CG017NICEguideline.

Nishida T, Hirota S, Yanagisawa A, *et al.* Clinical practice guidelines for gastrointestinal stromal tumor (GIST) in Japan: English Version. *Int J Clin Oncol* 2008;**13**:416–430.

Noda N, Sasako M, Yamaguchi N, Nakanishi Y. Ignoring small lymph nodes can be a major cause of staging error in gastric cancer. *Br J Surg* 1998;**85**:831–834.

Norton JA. Surgical treatment and prognosis of gastrinoma. *Best Pract Res Clin Gastroenterol* 2005;**19**:799–805.

Perini RF, Camara PRS, Ferraz JGP. Pathogenesis of portal hypertensive gastropathy: translating basic research into clinical practice. *Nat Rev Gastroenterol Hepatol* 2009;**6**:150–158.

Sato C, Naganawa S, Kumada H, *et al*. MR imaging of gastric cancer in vitro: accuracy of invasion depth diagnosis. *Eur Radiol* 2004;**14**:1543–1549.

Schmulewitz N, Baillie J. Dieulafoy lesions: a review of 6 years of experience at a tertiary referral center. *Am J Gastroenterol* 2001;**96**:1688–1694.

Wu CW, Hsiung CA, Lo SS, *et al*. Nodal dissection for patients with gastric cancer: a randomised controlled trial. *Lancet Oncol* 2006;**7**:309–315.

Zollinger RM, Ellison EH. Primary peptic ulcerations of the jejunum associated with islet cell tumors of the pancreas. *Ann Surg* 1955;**142**:709–723.

CHAPTER 24

Disorders of the liver

FRANCESCO M. POLIGNANO, JOHN F. DILLON, IAN A. ZEALLEY, AARON J. QUYN, AILEEN SMITH AND SIR ALFRED CUSCHIERI

Anatomy of the liver

The liver is the largest organ in the body, with a weight varying from 1200 to 1600 g. It arises from the foregut endoderm as a diverticulum which extends into the septum transversum and connects with the vitelline veins of the yolk sac. The caudal section of the hepatic anlage ultimately forms the biliary tract and gallbladder while the cephalic section forms the hepatic parenchyma. The vitelline veins form the portal and hepatic veins. The left umbilical vein persists as the ductus venosum and diverts oxygenated blood from the placenta around the liver directly into the inferior vena cava. After birth the vestigial ligamentum venosum runs in the free edge of the falciform ligament (round ligament, ligamentum teres). It may recanalize in patients with portal venous hypertension or can be used after dilatation for exchange blood transfusion or to permit radiological investigation of the portal venous system.

Morphology and topographical anatomy of the liver surfaces

The liver can be regarded as a wedge with rounded edges tapering to the left. It has three surfaces: anterosuperior, inferior and posterior. The anterosuperior surface is marked by the umbilical fissure in the depths of which (recessus of Rex) is inserted the round ligament (obliterated umbilical vein) attached to the cornu of the left portal vein. The umbilical fissure and the falciform ligament are the most conspicuous anatomical landmarks and divide the liver into right and left lobes. The posterior surface of the liver is largely formed of the bare area of the right lobe of the liver attached loosely to the diaphragm and retroperitoneum, and the caval canal that accommodates the retrohepatic vena cava and hepatic veins. The inferior surface of the liver is more complicated. The gallbladder is attached anteriorly some distance to the right of the umbilical fissure. The area between the gallbladder and the umbilical fissure is known as the quadrate lobe. Behind this and the neck of the gallbladder is the transverse hilar fissure that contains the main divisions of the portal vein, hepatic artery and common hepatic duct, and forms the posterior limit of the right lobe. The hepatic parenchyma separating the hilar fissure and the inferior surface of the left lobe from the vena cava forms the caudate lobe.

Relations

In terms of anatomical relations, the anterosuperior surface of the liver is in contact with the diaphragm and its upper margin reaches the level of the fourth interspace on the right and crosses the junction of the xiphisternum and sternum. Inferiorly the tip of the right liver reaches the costal margin, though Riedel's extension commonly extends below the costal margin and can reach the iliac crest. The gallbladder lies in the gallbladder fossa on the undersurface of the liver. There is a layer of fascia between the liver and the gallbladder (which must not be transgressed during cholecystectomy). Inferior relations of the right lobe include the upper pole of the right kidney and the right adrenal gland and more anteriorly the first part of the duodenum as it is overlapped by the gallbladder. The much thinner left lobe overlies the gastro-oesophageal junction. The lesser sac lies below the liver and behind the lesser omentum. It usually communicates with the rest of the peritoneal cavity through the foramen of Winslow lying behind the portal vein, hepatic artery and common bile duct. Occlusion of these structures (by finger, sling or vascular clamp) stops the arterial and portal venous inflow and this manoeuvre first described by Pringle is used to control bleeding during liver surgery and liver trauma.

Ligaments

The coronary and triangular ligaments suspend the liver from the diaphragm. The left triangular ligament is a thin peritoneal fold that is relatively avascular. Its two layers separate medially as the caval canal is reached: one sweeps back to the lesser sac and the other forms the left leaf of the falciform ligament. The coronary ligament is composed of two separate peritoneal folds (superior and inferior) between which lies the 'bare area' of the liver that is connected to the diaphragm by loose relatively avascular areolar tissue. The two leaves of the coronary ligaments join laterally, thus forming a V-shaped attachment to the diaphragm and retroperitoneum. The superior layer is an extension of the right leaf of the falciform ligament and attaches the superior surface to the diaphragm. The inferior leaf of the coronary ligament is a reflection of the peritoneum covering the right perinephric and adrenal region on to the inferior surface of the liver to the right of the infrahepatic vena cava.

Caval canal

This important region is often ignored in accounts of surgical anatomy. It is essentially a gutter in which lies the retrohepatic vena cava and the hepatic veins, all of which are enveloped in a loose fibrous meshwork rather than membranes. The caudate lobe separates the caval canal from the hilar fissure anteriorly and from the left lobe. The caval canal is best exposed during hepatectomy by a combined superior and inferior approach. Superiorly, the caval canal is covered by the diverging layers of the falciform ligament. When these are divided, a loose fibrous packing tissue envelops the vena cava and the right and left hepatic veins. Below, the caval canal is opened by division of the inferior leaf of the coronary ligament. The fibrous tissue covering the vena cava is loose in this region. A variable number of unnamed hepatic veins are encountered and these include veins to the caudate lobe. The retrohepatic vena cava also receives the phrenic veins and below these it is loosely attached to the retroperitoneum and, thus, provided the correct plane is identified, a sling can be passed around it and including the right and left hepatic veins in the immediate suprahepatic region.

Functional or segmental anatomy of the liver

The segmental anatomy of the liver on which modern hepatic surgery is based comes from the anatomical dissections performed by Rex (1888) and Cantlie (1898) more than 100 years ago and subsequently elaborated by Goldsmith and Woodburne (1957), Couinaud (1957), Healy and Schroy (1953) and Elias and Petty (1952).

In essence, the liver should be regarded as a paired organ (right and left livers) fused along a line extending from the middle of the gallbladder fossa anteriorly to the left edge of the suprahepatic inferior vena cava posteriorly. Within the liver this corresponds to a vertical plane (the main portal scissura or Cantlie's line) in which lies the middle hepatic vein. The right liver receives the right portal vein, right hepatic artery and right hepatic duct and the left liver the corresponding left portal vein, left hepatic artery and left hepatic duct.

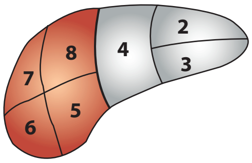
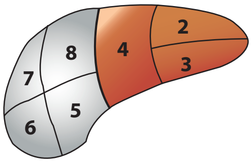
Considerable confusion is often caused by the discrepancy between surface and functional anatomy of the liver, with a portion of the right anatomical lobe (segment IV) effectively belonging to the left functional hemiliver. For this reason, it is advocated that the anatomical division in right and left lobe be left to the anatomists, whereas the terms right and left hemilivers are to be used in all clinical and surgical settings. A further element of confusion is generated by the subdivision of the two hemilivers in sectors and segments, which are effectively fully independent subunits, and by the different terminology used by the influential French school with its followers and the wider international community. A consensus document has recently been produced by the International Hepato-Pancreato-Biliary Association entitled 'IHPBA Brisbane 2000 Terminology of Liver Anatomy & Resections', which is reported in its entirety and whose use is now widely advocated (Figure 24.1).

Because of this segmental liver anatomy, it is possible to resect a single or several segments even in the liver that has been distorted by chronic hepatic disease. A careful identification of the vessels and ducts supplying each segment can be achieved by dissection above the portal hilum or within the liver parenchyma and each segmental or sectorial pedicle may be ligated separately prior to resection of the corresponding portion of liver realizing an 'anatomical resection'.

The right hemiliver

The right hemiliver, situated to the right of Cantlie's line, is further subdivided into two sectors by the vertical right portal scissura in which lies the right hepatic vein. This scissura, which has no surface markings, is a plane that subtends an angle of 40° with a coronal plane conducted through the inferior vena cava (horizontal plane) and on the liver surface corresponds to a line extending from the anterior edge of the liver (midway between its rightmost tip and the right border of the gallbladder fundus) to the confluence of the right hepatic vein with the vena cava. The right posterior sector is situated to the right of the right portal scissura (and of the right hepatic vein) while the right anterior sector (also known as the right paramedian sector) is to the left of it, between the right and middle hepatic veins. The two

1 *First-order division*

Anatomical term	Couinaud segments referred to	Term for surgical resection	Diagram (pertinent area is shaded)
Right hemiliver **OR** *Right liver*	*Sg 5–8(+/–Sg1)*	*Right hepatectomy* **OR** *Right hemihepatectomy* *(stipulate +/–segment 1)*	
Left hemiliver **OR** *Left liver*	*Sg 2–4 (+/–Sg1)*	*Left hepatectomy* **OR** *Left hemihepatectomy* *(stipulate +/–segment 1)*	

Border or watershed: The border or watershed of the first-order division which separates the two hemilivers is a plane which intersects the gallbladder fossa and the fossa for the IVC and is called the midplane of the liver.

Figure 24.1 International Hepato-Pancreato-Biliary Association Brisbane 2000 Terminology of Liver Anatomy and Resections. IVC, inferior vena cava.

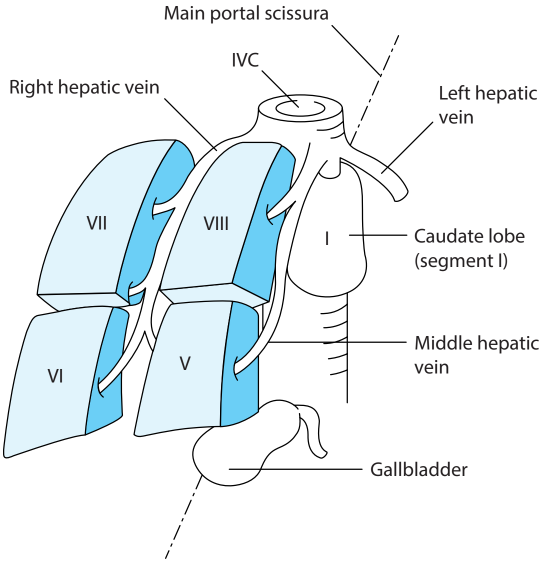


Figure 24.2 Schematic representation of the right liver. The main portal scissura (plane) containing the middle hepatic vein marks the territory between the two livers. It is also known as Cantlie's line and roughly makes an angle of 75° with the horizontal to the left, and extends from the middle of the gallbladder fossa to the left side of the vena cava. The right liver is divided into two sectors by the vertical right portal scissura containing the main trunk of the right hepatic vein. These two sectors are split by a horizontal plane through the right portal vein into two anteroinferior (V and VI) and two posterosuperior (VII and VIII) segments. Thus the right liver is smaller than the right lobe. Unfortunately, the vertical scissurae (planes) are undulating and not straight as shown in the drawing. This precludes a good correlation between radiological and anatomical segmentation (see text). IVC, inferior vena cava.

sectors are further divided in segments by a horizontal plane passing through the hilum of the liver (hilar plane); the anterior sector is therefore composed of segments V (anteroinferior) and VIII (anterosuperior) and the right posterior sector of segments VI (posteroinferior) and VII (posterosuperior). Each segment receives a portal pedicle (a segmental branch of the portal vein and a segmental branch of the hepatic artery) and is drained by a separate bile duct. Thus in the supine patient, segments V and VIII partially overlap segments VI and VII, respectively (Figure 24.2).

The left hemiliver

The left hemiliver is divided by the umbilical fissure in the left lateral sector (situated to the left of it) and left medial sector (segment IV also known as the left paramedian sector). The umbilical fissure is easily recognized on the liver surface by the presence of the round ligament (ligamentum teres) and of the falciform ligament. The left portal scissura, in which lies the left hepatic vein, has no surface marking and is said to run from the left margin of the suprahepatic inferior vena cava to the mid-point of the anteroinferior margin of the left lobe of liver. It divides the left lateral sector into an anterior segment (segment III) and a posterior segment (segment II) (Figure 24.3).

Caudate (dorsal, Spigel) lobe

Although this is customarily labelled as segment I, it is really a separate 'liver' because it has its own hepatic veins (Spigelian veins) and bile ducts, although it receives portal and arterial branches from both right and left sides. The caudate lobe is situated behind the hilar fissure and embraces the vena cava from the left forming an L-shaped structure with the horizontal limb separating the hilar fissure from the vena cava and the vertical limb the left liver from the vena cava. Its uppermost part lies posteriorly to the confluence of the hepatic veins and anteriorly to the inferior vena cava.

Main hepatic veins

Each numbered segment contributes hepatic veins that coalesce to form the main venous drainage of the livers and lie between the segments. There are three main veins of surgical importance:

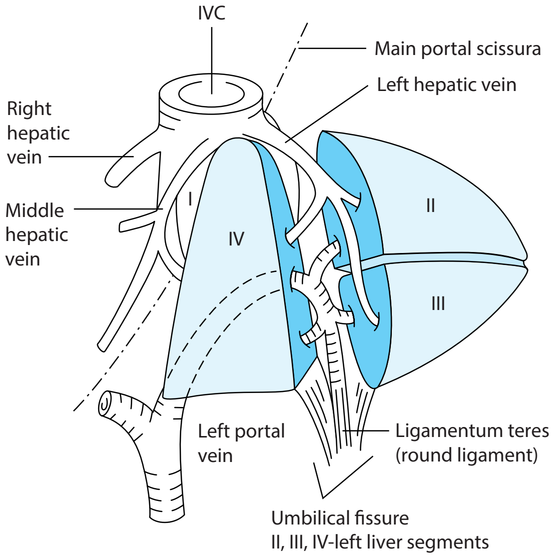


Figure 24.3 Schematic representation of the left hemiliver that is split by the umbilical fissure and by the left portal scissura into segment IV medially and segments II and III laterally. Thus the left liver is larger than the left lobe. IVC, inferior vena cava.

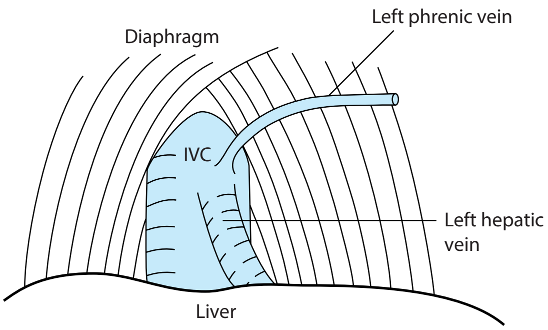


Figure 24.4 Schematic representation of the left hepatic vein. IVC, inferior vena cava.

the right hepatic vein drains segments V–VIII by a short vessel directly into the suprahepatic vena cava; the middle hepatic vein drains from both livers (segments IV and V–VIII) and empties either directly into the vena cava in isolation or with a common trunk with the left hepatic vein. The latter vein drains segments II and III. Segment I, the caudate lobe, drains by one or more small hepatic veins directly into the retrohepatic vena cava.

In the majority of patients both the right hepatic and the left hepatic veins can be identified and secured extrahepatically within the caval canal, but this manoeuvre is often hazardous for the middle hepatic vein. The left hepatic vein is exposed by complete division of the left triangular ligament. It slopes down from the vena cava to the liver so that its posterior wall is in contact with the vena cava and separated from it by loose fibrous tissue (Figure 24.4). The right hepatic vein is exposed only after the right lobe is completely mobilized and dislocated to the left. It runs directly posteriorly from the liver in intimate contact with the vena cava before entering it more posteriorly (Figure 24.5). Again only a loose layer of fibrous tissue separates the medial wall of the right hepatic vein from the vena cava but the fibrous tissue on its outer surface is usually firm and thickened.

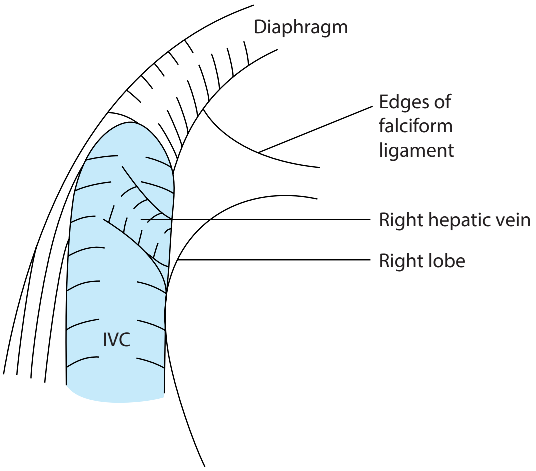


Figure 24.5 Schematic representation of the right hepatic vein after dislocation of the right lobe of the liver. IVC, inferior vena cava.

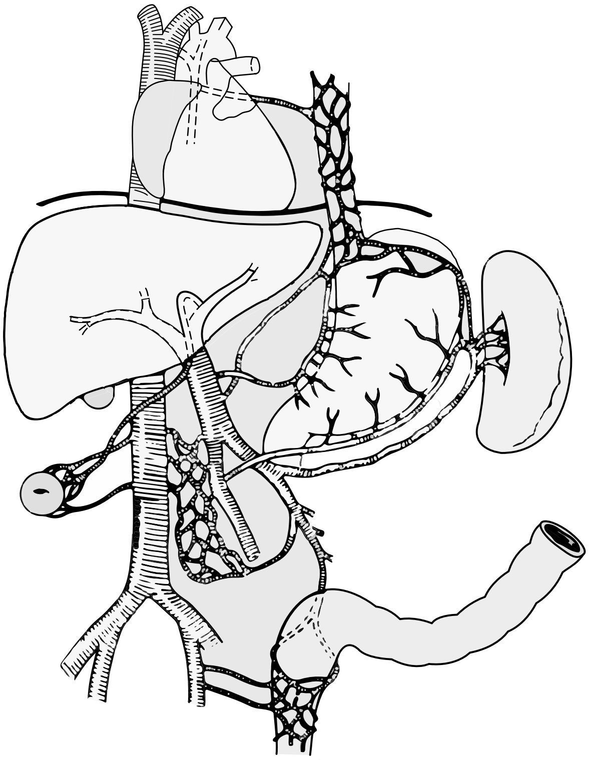


Figure 24.6 This shows the portal venous drainage from the gastrointestinal tract and demonstrates the major anastomotic sites between the portal and systemic systems: the cardio-oesophageal junction leading up to the azygos system, the retroperitoneum, the umbilicus and the inferior rectal plexus.

Portal hilum

In the portal hilum, the portal vein that has formed behind the head of the pancreas with the confluence of the splenic and superior mesenteric veins passes along the edge of the lesser omentum for 7.5 cm. It receives branches from the pylorus and the important left coronary vein from the cardio-oesophageal region (Figure 24.6). The portal vein provides the final conduit for the venous return from the gastrointestinal tract and its usual distribution to the various functional segments is shown in Figure 24.7. In the axial plane, the two main branches of the

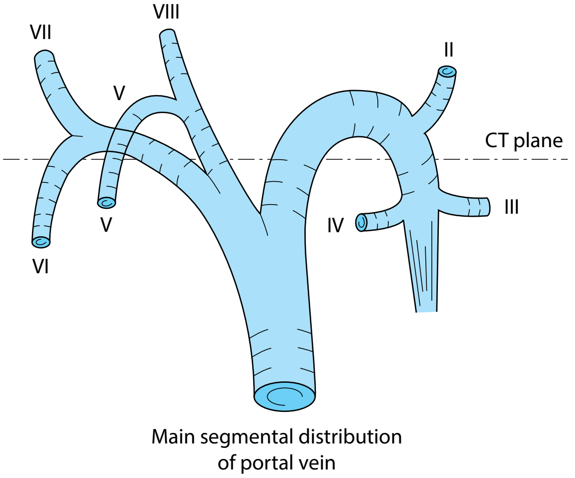


Figure 24.7 Schematic representation of the usual intrahepatic distribution of the portal vein. Above the transverse line (representing the transverse scissura marking the plane of the right and left portal veins) the upper branches supply segments VII, VIII, I, II and the lower branches VI, V, IV, III.

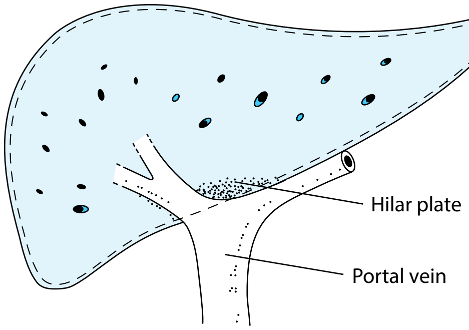


Figure 24.8 To the undersurface of the liver, the bifurcation of the major hilar structures is secured by a dense fibrous sheath termed the hilar plate. After incising this structure the vessels to the left liver segments run for 1.5 cm before dividing into segmental branches.

portal vein lie in the same plane (sometimes referred to as the transverse scissura). From right to left, segments VI, V, IV and III are below this plane, whereas segments VII, VIII, I (caudate) and II are above it.

There are major anastomotic sites between the portal and systemic systems that open up in the presence of occlusion/obstruction to portal blood flow to the liver:

- the cardio-oesophageal junction – left gastric (coronary) vein to the azygos system
- communications with the retroperitoneal veins of Sappey
- umbilicus – recanalized left umbilical vein to abdominal parietal veins – caput medusae
- communications with the inferior rectal plexus (Figure 24.6).

The common hepatic duct draining both livers passes in front of and to the right of the portal vein and receives the cystic duct at a variable point of its course to form the common bile duct. The common hepatic artery runs to the left of the common bile duct giving off the main cystic artery and branches to the common bile duct prior to division into right and left branches.

An understanding of the point of the division of the structures in the portal hilum is essential for the surgeon. The vasculobiliary sheath, described originally by Walaeus in 1640 and commonly referred to as the Glissonian sheath, surrounds the main vessels and ducts following them well into the depth of the parenchyma and it is still recognizable around the tertiary ramifications (segmental pedicles). While the portal vein is loosely enclosed, the bile duct and hepatic artery are firmly adherent to the Glissonian sheath. The upper surface of the sheath which is in contact with the liver parenchyma thickens to become the hilar plate. This structure can be released from the liver surface because there are no branches along it and permits the surgeon to isolate the bifurcation of the hepatic pedicle to the right and left hemilivers extrahepatically but *en bloc* (the three elements together, portal vein, hepatic artery and bile duct) and to proceed to liver resection more safely with an extraglissonian approach (Figure 24.8).

Techniques of anatomical liver resection and intraoperative ultrasound

Different approaches to liver resection surgery have been developed during the years, with the most influential contributions coming from the French (Couinaud, Lotart-Jacob and Bismuth) and Vietnamese (Ton That Tung) schools. More recently the wide adoption of intraoperative ultrasound during liver surgery has facilitated the recognition of intrahepatic structures, like the hepatic veins and the second- and third-order portal pedicles, and has contributed enormously to the spreading of more complex resections like the mesohepatectomy (right anterior and left median sectoriectomy, equivalent to a resection of segments V, VIII and IV) or to anatomical resections of the caudate lobe, of segment VIII or of segment VII in isolation (Figure 24.9).

The French school has popularized a technique whereby major hepatectomies are performed ‘with preliminary vascular division’, where the elements of the vascular pedicle to the liver being resected are dissected out and divided extrahepatically. The relevant hepatic vein is also isolated and divided prior to commencing the transection of the liver parenchyma. In a right hepatectomy, for example, this would typically involve a meticulous dissection within the hepatoduodenal ligament to identify and divide individually the elements of the right glissonian pedicle (right hepatic artery, the right branch of the portal vein). The right hepatic vein would then be dissected out and divided prior to commencing the hepatotomy. The right hepatic duct is usually left intact and divided at the very end of the operation once the frequent anatomical variants of the biliary tree have been ruled out. With this technique, a clear ischaemic demarcation of the liver being resected is determined, which facilitates the subsequent hepatotomy and allows a reduction in intraoperative blood loss resulting from the preventive devascularization of the prospective specimen. However, the risks posed by the difficult dissection of the hepatic veins and of the three glissonian elements individually are well recognized and complications like major haemorrhages from the inferior vena cava or hepatic veins and unintentional devascularization or interruption of bile ducts to the prospective liver remnant have been widely reported, the latter particularly in the presence of anatomical variants. Furthermore, with this

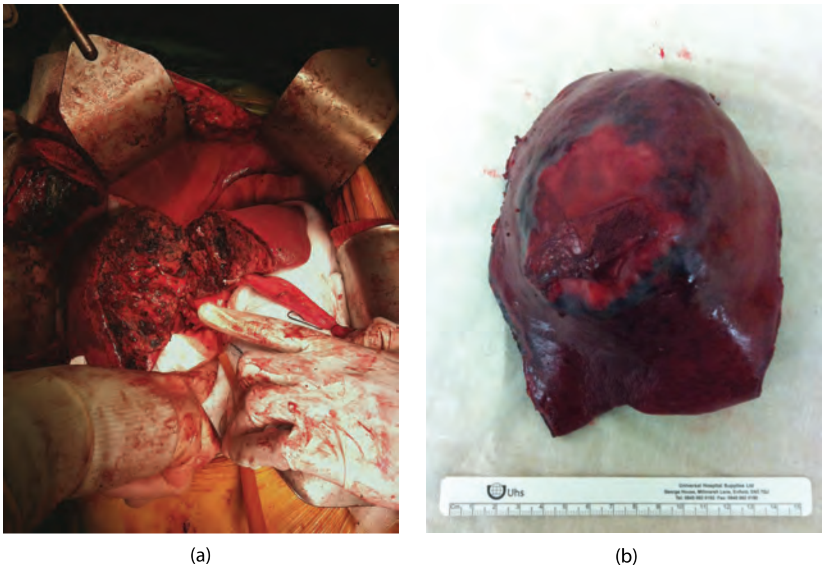

(a) (b)

Figure 24.9 (a) Surgical view after a mesohepatectomy (resection of segments IV, V, VIII) for large colorectal cancer metastasis centred in segment IV–VIII and corresponding surgical specimen. (b) A small section of diaphragm is visible attached to the tumour surface.

technique, the anatomical resection of sectors or liver segments in isolation is more challenging, as the secondary and tertiary glissonian pedicles are located very deep in the liver and are very difficult if not impossible to access without a certain degree of dissection within the liver parenchyma.

A recent modification of this technique is the approach to liver resection 'with selective clamping', where structures at the hepatic hilum are still dissected intraglissonially according to Lotart Jacob but are simply clamped and not irreversibly divided, thus allowing the surgeon more time to interpret the anatomy variants while preserving the benefits of preventive devascularization of the prospective specimen and of its ischaemic demarcation.

In contrast to this technique, the Vietnamese surgeon Ton That Tung proposed in the late 1960s an approach to liver surgery 'with primary parenchymal transection', based on in-depth knowledge of the intrahepatic anatomy and corresponding surface landmarks and on the observation that, even though variations of the extrahepatic arterial, portal and biliary anatomy are common, the intrahepatic distribution of these structures and their relation with specific surface landmarks is surprisingly consistent. With this technique the hepatotomy is started at the outset along the line of a scissura and the glissonian pedicle to the prospective specimen is dissected out and divided within the depth of the liver parenchyma, usually preserving the integrity of its glissonian sheet. The relevant hepatic vein is also reached and divided inside the liver parenchyma and no attempt is made at extrahepatic dissection of it. The main advantage of this approach lies in its safety, because by approaching the glissonian pedicle inside the liver substance, well above the hilum, and by preserving the integrity of its sheet, all possible anatomical variants are easily circumvented and the difficult and hazardous intraglissonian dissection at the hilum can be avoided. Fear of a more significant blood loss or a more prolonged clamping time of the hepatoduodenal ligament (Pringle manoeuvre) and the need of a rapid execution were often listed as downsides of the Ton That Tung technique and have ultimately hampered its spreading in Europe and in the Western world for a long time. Poor understanding of intrahepatic anatomy in the Western world has also been a contributing factor. Nowadays, however, the availability of faster and more haemostatic devices for the transection of the liver (CUSA and radiofrequency devices like the Tissuelink), and of intraoperative ultrasound to guide the parenchymal transection, has determined a resurgence of interest around this technique and its principles, in its original description or in modified forms.

The extrahepatic extraglissonian approach has also been described in combination with the conventional extrahepatic approach to the relevant hepatic vein during a right hepatectomy or posterior sectoriectomy and during mesohepatectomy. It has also been proposed for resections '*à la demande*' where strict respect of the segmental borders is not possible.

Whichever the technique employed, intraoperative ultrasound has most certainly become mandatory during liver surgery and has greatly increased its safety. It has allowed liver surgeons to circumvent the lack of surface landmarks by direct visualization of hepatic veins or pedicles within the parenchyma. It also permits direct visualization of the deep tumour margins, which is essential to achieve a radical resection. In expert hands, the identification of anatomical variants can also be achieved by intraoperative ultrasound and the identification of unrecognized pathologies or additional liver lesions, which can ultimately lead to a change of surgical strategy.

Hepatic architecture

Conventional morphology considers that the liver is composed of pyramidal lobules based on a central vein and surrounded on the periphery by portal trunks with terminal radicles of bile duct, portal vein and hepatic artery (Figure 24.10). The two vascular systems of the central vein and portal tract lie on planes at right angles to one another and never interdigitate. Thus the sinusoids

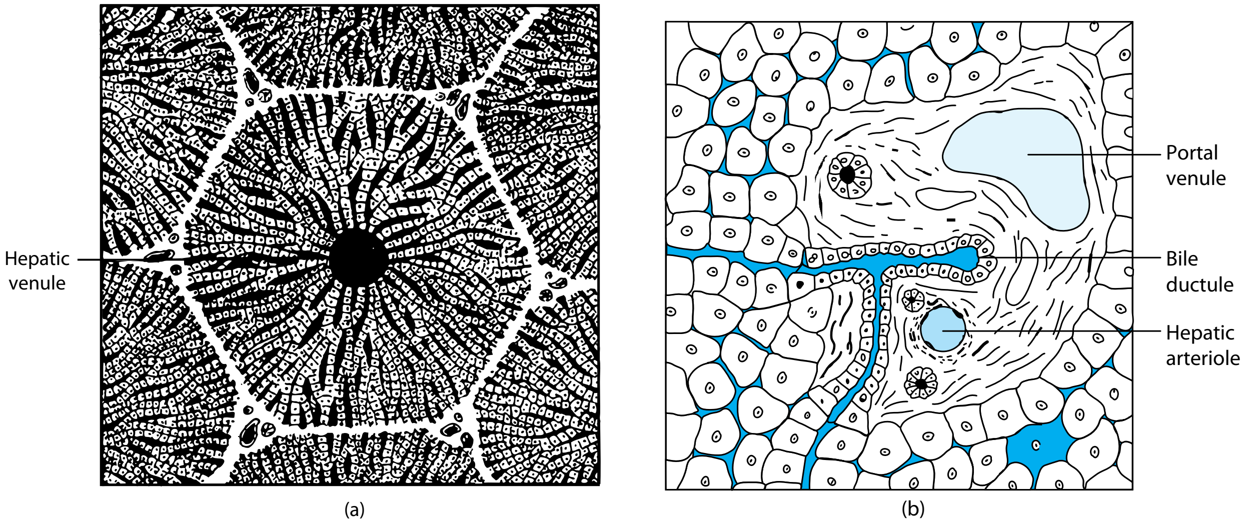


Figure. 24.10 (a) This demonstrates the normal liver architecture in which a hepatic plate apparently surrounds a hepatic vein through which blood from the hepatic sinusoids drains. (b) In fact it is more appropriate to centre the hepatic plate on the portal tract which contains the elements of the bile ducts into which bile can directly drain from the hepatocytes and branches of the portal vein and hepatic artery together with connective tissue stroma. Zones closest to this portal tract (zone 1) are protected in most circumstances from liver damage. Liver cells surrounding the hepatic vein (zone 3) are more susceptible to all forms of hepatic damage. Regenerative nodules are normally centred on a portal tract.

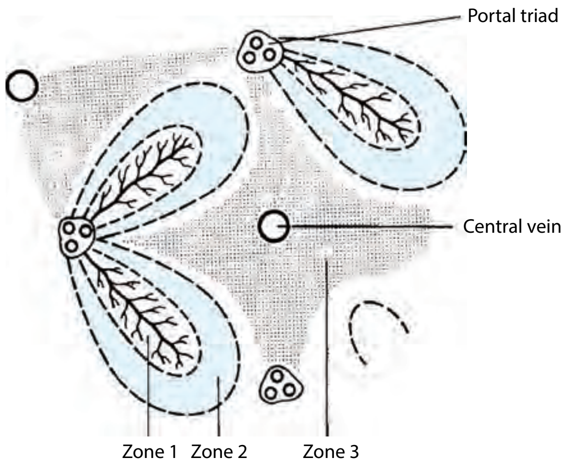


Figure 24.11 According to Rappaport, there are three zones of liver parenchyma. Zone 1 adjacent to the portal triad is the best vascularized and least susceptible to injury. Zone 3 is adjacent to the central vein and most susceptible to injury.

are arranged perpendicular to the planes of the central veins and portal blood passes to the central vein along a pressure gradient. The walls of the sinusoids are composed of endothelial and phagocytic cells, termed Kupffer cells. Between the hepatocytes and Kupffer cells is the space of Disse. Bile canaliculi are shown to be channels or grooves in the hepatocyte surface, lined by microvilli. The network of canaliculi drains the liver lobules into the terminal bile ducts.

It may help in the understanding of liver injury and its consequences to view the liver morphology somewhat differently. The concept of Rappaport is to regard the liver as a series of acini supplied by a portal triad of structures (Figure 24.11). Three zones of sinusoids are envisaged in which the peripheral zone of the acini (zone 3) is damaged more severely in any form of injury. Adjacent forms of injury may coalesce to form areas of bridging necrosis and, later, fibrosis, producing the common pattern of postsinusoidal block; zones 1 and 2 may form the nidus of surviving cells which then regenerate in nodular form.

Liver function tests and testing liver function

The analytes commonly called liver function tests (LFTs), alanine amino transaminase (ALT), aspartate transaminase (AST), alkaline phosphate and -glutamyltransferase, do not give any information about liver function and are biomarkers of liver damage. The other biochemical analyses commonly performed as part of the LFTs group are albumin and bilirubin: these do give some indication of liver function but with important caveats. Other biochemical abnormalities include elevations of acute phase reactant proteins ($_2$-macroglobulin, C-reactive protein, etc.). Other changes include lowering of serum proteins, release of integral proteins, raised immunoglobulins and the appearance of abnormal protein antigens.

Liver function tests

In hepatocellular injury, integral cellular enzymes are released into the circulation. These include AST and ALT. The levels of both transaminases in the serum reflect the severity of the ongoing liver cell damage and necrosis. Minor elevations are encountered in cholestasis and chronic liver disease, whereas substantial serum levels (>400 units) are encountered in acute hepatitis with liver cell necrosis from any cause; acute cholangitis can also be associated with a very elevated ALT, sometimes causing diagnostic confusion. The ALT is more specific for liver tissue, and is often the only transaminase routinely measured; however, the ratio of ALT–AST can give additional information about hepatic fibrosis or non-alcoholic fatty liver disease (NAFLD).

The group of isoenzymes responsible for the hydrolysis of phosphate esters in an alkaline medium are termed alkaline phosphatases. These enzymes are found in the liver, biliary tract, bone, intestine, kidney and placenta. The various isoenzymes can now be differentiated by special immunoassays establishing the exact tissue origin of the circulating enzyme. In cholestasis and obstructive jaundice, the conjugated hyperbilirubinaemia is usually accompanied by a significant elevation of the serum alkaline phosphatase activity due to impaired excretion of the enzyme in the bile and hepatic induction (overproduction). Other (alkaline phosphatase type) enzymes associated with the biliary tract such as 5-nucleotidase and -glutamyl transpeptidase (-GT) exhibit a parallel elevation. The -GT is highly inducible in hepatocytes, by drugs and alcohol so may be disproportionately elevated in some patients. It should not be regarded as diagnostic in isolation; it should be interpreted in the context of the rest of the LFTs and the clinical situation. The biochemical picture in patients with diffuse infiltrative lesions of the liver or secondary deposits is often of a marked elevation of the bile ductular enzymes (alkaline phosphatase, -GT) out of proportion to the small rise in the serum bilirubin; this observation is a pointer to performing further imaging; the absence of these changes does not indicate the liver is uninvolved in infiltrative disease.

Albumin

The serum albumin is often low in patients with chronic liver disease and the hypoalbuminaemia is associated with a complementary rise in the other serum proteins, particularly globulins (altered albumin–globulin ratio). The fall in the serum albumin is often used in the grading of the severity of hepatic decompensation, e.g. Child–Pugh's classification. When severe, the hypoalbuminaemia is usually accompanied by fluid and salt retention, but reduced plasma oncotic pressure is only one of several interrelated factors responsible for ascites in patients with chronic liver disease. The exact mechanism (defects in transcription, mRNA processing or translation, defective export) for the depressed albumin synthesis in patients with hepatic fibrosis or cirrhosis is not known. Paradoxically, in some patients with cirrhosis, albumin synthesis by the liver is actually increased, but often not enough to compensate for the expansion in extracellular fluid.

Immunoglobulins

Marked elevation of the serum immunoglobulin (Ig)G is encountered in patients with cirrhosis and autoimmune hepatitis, but is non-specific and more a marker of activity than a diagnostic aid. Serum IgA, which is normally secreted in the bile, is also elevated in these conditions but to a lesser extent and may also be elevated in alcohol-related liver disease, but with diagnostically poor sensitivity and specificity. Primary biliary cirrhosis is associated with elevation of the serum IgM and the presence of antimitochondrial antibodies.

Biomarker proteins

The only commonly available biomarker is α-fetoprotein (AFP), which is detected in the serum of about 60% cases of hepatocellular carcinoma (HCC). AFP is commonly used in the surveillance of patients at increased risk of HCC, but its use is now controversial, with guidelines advocating its use only in combination with hepatic ultrasound examination and some evidence suggesting AFP has no added benefit to ultrasound. An AFP level of over 100 kU/L is highly suspicious of HCC, and a level between 7 and 100 kU/L may be indicative of a small HCC (<5.0 cm) but is also encountered in benign liver disease associated with necrosis and regeneration, e.g. hepatitis C. A rising level of AFP even if below the diagnostic range is indicative of HCC and should provoke a search using hepatic CT or MRI. AFP is most useful in monitoring patients with HCC, which is known to produce AFP. Persistent elevation after resection of HCC indicates residual disease and re-elevation after initial normalization is suggestive of recurrence. A correlation between serum level of AFP, tumour size and survival has been established by a number of studies. Thus patients with marked preoperative elevation have a poor survival after surgery. The well-differentiated variant of HCC, fibrolamellar carcinoma, does not cause an elevation of AFP and this oncofetal antigen is not expressed by cholangiocarcinoma.

Measuring liver volume

Advances in imaging and software have allowed the development of *in vivo* imaging of the liver. Three-dimensional models of the liver can be constructed from CT or MRI scans. The volume of liver can then be calculated based on known separation of image slices combined with planar mapping of cross-sectional areas. In addition, these three-dimensional models can be simulated to map the effects of surgery by performing virtual hepatic resections. Studies have shown this to be of value as a predictor of hepatic dysfunction after major liver resection.

Testing liver function 'hepatic functional reserve'

In assessing a patient with a liver condition requiring surgical intervention and/or a patient with underlying parenchymal liver who requires major surgery, the usual decision-making process about the correct intervention has to take into account the influence of the liver on the prognosis of the condition. In clinical practice assessment of the hepatic reserve is based predominantly on standard 'liver function tests' together with a clinical assessment of nutritional state, muscle wasting and degree of sodium retention (based on peripheral oedema and ascites). Prothrombin time is a sensitive test of hepatic synthesis of vitamin K-dependent factors that is impaired in jaundice and hepatocellular disease. In addition to the level of prothrombin activity, the response to a parenteral dose of vitamin K analogue is a most useful test of hepatocyte function. Thus restoration to normal within 24–48 hours is expected in patients with obstructive jaundice without significant impairment of hepatocyte function. Failing to do this shows significant impairment of synthetic liver function. These are often combined into scoring schemes, the most commonly used are the Child–Pugh score/grading (modified from the initial Child–Turcotte). The Child's score is based on bilirubin, albumin, prothrombin time, presence of ascites and presence of encephalopathy (Tables 24.1 and 24.2). The Child's score has

Table 24.1 Child–Pugh classification of disease severity in cirrhotic patients

Parameter	A	B	C
Albumin (g/dL)	>3.5	3.0–3.5	<3.0
Bilirubin (mol/L)	<25	25–40	>40
Prothrombin (s) >normal	<4	4–6	>6*
Prothrombin (%)	>64	40–65	<40
Ascites	None	Controlled	Refractory
Encephalopathy	None	Minimal	Advanced

* In the original Child–Turcotte classification nutrition was used but this has been substituted with prothrombin activity. The bilirubin has to be adjusted for patients with primary biliary cirrhosis (A, 5–7; B, 8–10; C, 11 or more).

Table 24.2 Paul Brousse hospital classification system

Parameter	Number of criteria
Albuminaemia <3.0 g/100 mL	1
Hyperbilirubinaemia >30 mol/L	1
Encephalopathy	1
Clinical ascites	1
Coagulation factors II and V 40–60%	1
Coagulation factors II and V <40%	2

A, none of the criteria.
B, 1 or 2 criteria.
C, 3 or more criteria.

some utility at identifying those patients with good prognosis and those with very bad prognosis but many patients that it is applied to have an intermediate B class, which has a much wider range of prognosis in studies so is of limited usefulness to the individual patient. The predictive tools developed for allocation of transplant organs, i.e. MELD or UKELD, are increasingly being used as predictors of poor outcome for non-transplant surgical intervention; they are promising in this role but require further validation.

There are in addition tests that give an assessment of the functional capacity (reserve) of the liver. These are not often used clinically except in specialist hepatobiliary centres and in clinical research studies. Some are tests of microsomal function (e.g. antipyrine clearance, caffeine clearance), others of cytosolic function (e.g. galactose elimination capacity) or hepatic perfusion and excretion [e.g. indocyanine green (ICG) clearance] or synthesis (albumin, urea and prothrombin). In practice, the most commonly used are ICG clearance and the prothrombin activity after administration of vitamin K.

Indocyanine green clearance

This is presently considered to be the best test of hepatic functional reserve. Previously it had been most commonly used to select patients for hepatic resection. More recently, ICG clearance has been reported to be beneficial in liver transplantation – selection of patients and prediction of graft viability. Hepatic clearance of ICG is dependent on many factors, the most important of which are hepatic blood flow, binding to plasma proteins, transport across the sinusoidal–hepatocyte membrane, intracellular transport, export across the canalicular membrane and bile flow. In clinical practice, hepatic clearance is derived from the plasma ICG clearance curve. The test is carried out after an overnight fast when ICG (0.5 g/kg body weight) is injected intravenously, usually in the antecubital vein. Blood samples are taken from the contralateral antecubital vein before administration (plasma blank) and at 5, 10, 15 and 20 minutes after injection of ICG. Most consider that a retention of greater than 14% at 15 minutes precludes major resection as it indicates a significant reduction in the hepatic reserve.

However, good survival has been reported in patients having an ICG retention greater than 14% provided the resection is carried out expertly and is not accompanied by significant blood loss. The limitation of this technique is its dependence on hepatic blood. In many patients in whom resection is being considered and there is concern about residual liver function, there will be underlying cirrhosis, which makes the ICG clearance unreliable, depending on the degree of portosystemic shunting.

Imaging of the liver

General considerations

A wide range of diagnostic imaging investigations are useful for diagnosis and assessment of diffuse liver disease, and for the identification and characterization of focal liver lesions. However it is important to remember that imaging is an adjunct to, and not a replacement for, clinical evaluation. The choice of imaging investigation and its interpretation must be guided by factors unique to each patient. Every patient presents a different combination of age, sex, risk factors, comorbidities and expectations, and imaging tests should be chosen and interpreted in the light of these. Although numerous 'protocols' for imaging exist, these are no substitute for comprehensive evaluation of individual patients and consideration of the diagnostic requirements in each unique case.

When considering a patient who is clinically suspected of having a diffuse liver disease such as cirrhosis there are a variety of imaging findings which may assist the diagnosis and go some way to determining the severity of the disease. However, when considering the various clinical, laboratory and imaging findings the clinician must consider whether for *this individual patient* they are sufficiently confident in the diagnosis to manage the patient effectively. They must always consider whether a biopsy of the liver tissue will or will not substantially improve the confidence and appropriateness of patient management.

Similar considerations apply when considering a patient in whom focal liver lesions have been identified or are suspected. There are a variety of imaging findings which may assist the identification and characterization of focal liver lesions, but there are very few such lesions which present an absolutely diagnostic appearance on which a clinician can rely with complete confidence. Lesions which may be diagnosed with absolute confidence using non-invasive imaging techniques are simple liver cysts and cavernous haemangiomas. All other focal liver lesions may present a range of findings on imaging tests and the best that may be achieved is to suspect, but never attain, complete confidence in diagnosis. An important example is the overlap which exists between the imaging findings that may be presented by tumours of hepatocyte origin at the benign and malignant ends of the spectrum. For instance a HCC lesion may present imaging features which are identical to those typically considered characteristic for a benign lesion such as hepatic adenoma or focal nodular hyperplasia (FNH) (Figure 24.12). As with diffuse liver disease, the clinician must consider how the various clinical, laboratory and imaging findings combine to influence the management of the *individual patient*, who has their own unique combination of age, risk factors and comorbidities. In essence, the question which the clinician must ask themselves is whether, for this *individual patient*, there is an important possibility that they may be undertreating on the basis of non-invasive tests alone, or is further investigation with biopsy or laparoscopy warranted.

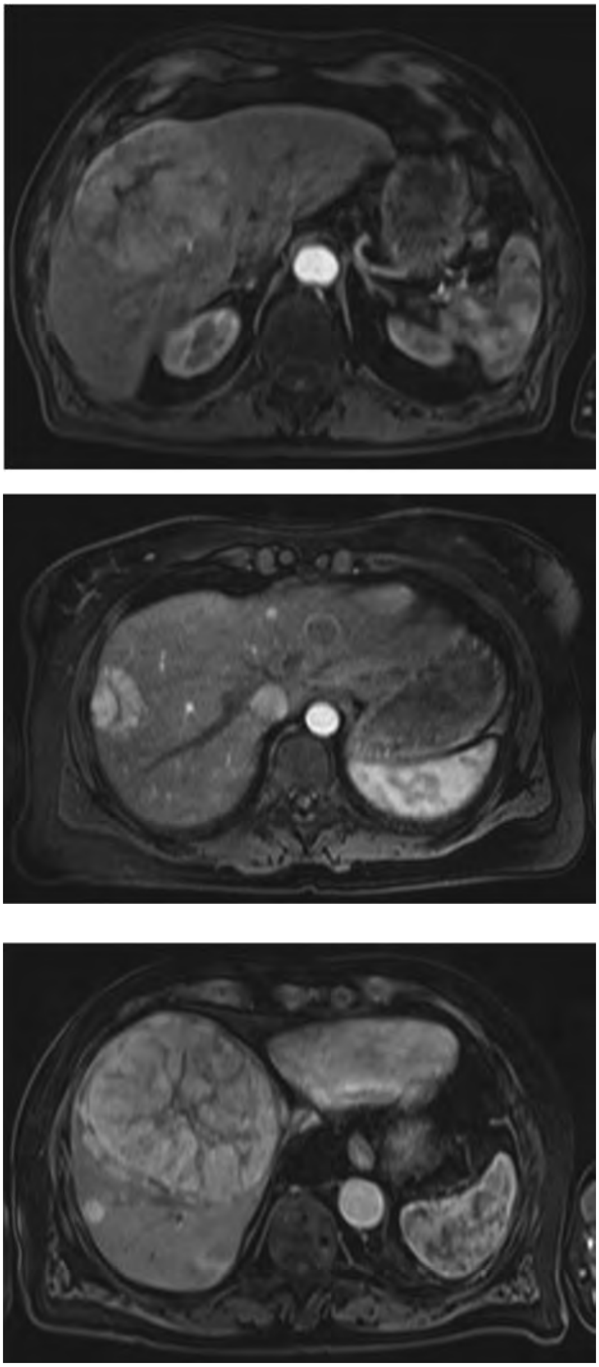


Figure 24.12 Axial MR images of three different focal liver lesions each exhibiting similar characteristics including enhancement during the arterial phase following intravenous gadolinium contrast administration and central 'scars'. These features are typical for focal nodular hyperplasia (FNH) but (a) is a hepatocellular carcinoma, (b) is a FNH lesion, and (c) is a fibrolamellar hepatocellular carcinoma.

Different imaging tests have different strengths and weaknesses. For any specific clinical scenario there is usually an optimal test. In addition there are often a variety of different ways in which an imaging test may be performed. Effective communication between clinician and radiologist is essential in order to maximize the diagnostic utility of imaging tests for the patient. Performing the right test in the right way minimizes the potential for 'false reassurance', when an inappropriate test fails to identify pathology which is in fact present, and also minimizes the potential for identification of 'pseudolesions' and irrelevant incidentalomas which may cause unnecessary and unwarranted distress to the patient.

Plain radiography

Plain abdominal radiographs may give some indication of the size of the liver but they rarely assist in the confident diagnosis or exclusion of liver disease. Mature hydatid cysts may be identified by virtue of calcification within the cysts wall, and liver abscesses may contain sufficient gas to be identified on a plain radiograph.

Ultrasonography

Because it does not involve ionizing radiation and is relatively inexpensive, ultrasonography is the primary investigation in patients with suspected biliary tract pain, cholestasis and non-dyspeptic upper abdominal pain. The normal liver is seen as a large homogeneous mid-grey structure in which the portal vein, hepatic veins and inferior vena cava are well shown. In patients with cholestasis the observation of dilated intrahepatic bile ducts throughout the liver indicates the presence of a large duct obstruction and occasionally the cause for this obstruction may be seen downstream in the biliary tree. Gallbladder disease is particularly well demonstrated, and gallstones and cholecystitis may be diagnosed reliably in most cases.

Ultrasound may also be employed to demonstrate the portal venous system. The patency of the portal vein and other major vascular channels, including surgical shunts, is usually well demonstrated with combined greyscale and Doppler ultrasound scanning ('duplex' scanning). The colour Doppler facility demonstrates the direction of flow.

A specialized modification of liver ultrasound involves the intravenous administration of contrast agents which enhance the image contrast between normal and abnormal liver tissue. These agents are stabilized forms of microscopic bubbles and may be employed either to increase the sensitivity of ultrasound for the detection of focal liver lesions or to aid characterization of focal liver lesions once found.

When screening the liver for focal lesions, ultrasound may identify lesions as small as 1 cm. Echo-poor liver cysts and profoundly echogenic cavernous haemangiomas are particularly well demonstrated. Liver neoplasms including primary liver tumours (benign and malignant) and metastases may also be seen. Used as a screening test in patients at high risk for development of HCC, for instance those with chronic hepatitis B or C infection, ultrasound has been demonstrated to identify patients with small, curable tumours in whom prolonged survival has been reported.

Once focal liver lesions have been identified (by whatever means), ultrasound may be used to follow up known lesions in order to assess for stability, growth or regression which may help to clarify diagnosis and influence management.

Diffuse liver disease may be inferred from ultrasound findings. In cirrhosis the characteristic features are generalized hyperechogenicity which is typically irregular and 'grainy'. The degree of fibrosis may be estimated non-invasively by means of ultrasound elastography but this application is still at an experimental stage. In diffuse hepatic steatosis a more diffuse generalized hyperechogenicity is typically observed.

Intraoperative ultrasound screening of the liver is an excellent adjunct to operative evaluation of the liver and may be performed during laparoscopic or open surgery. The ultrasound probes employed are of the curvilinear or linear array configuration with high frequencies in the 7.5–10 MHz range providing very high definition, particularly in the near field (2.0–3.0 cm depth). These provide the best achievable demonstration of the anatomy of vascular structures and the boundaries of liver tumours. They may also identify previously undetected lesions which may influence the planning or progress of hepatic resection. One of the most useful applications of intraoperative ultrasound is to take advantage of its unparalleled resolution to identify previously undetected foci of malignancy in the residual liver remnant following curative hepatic resection. Identification and characterization of such lesions (sometimes assisted by intraoperative ultrasound-guided biopsy) may allow the surgeon to avoid potentially high-morbidity surgery in patients who will not benefit. Intraoperative ultrasound may also be used to clarify the relationship between tumour and vascular structures in order to confirm the feasibility of planned operative procedures before embarking on dissection.

CT scanning

CT scanning is an extremely useful modality for diagnosis, follow-up and treatment planning in liver disease. Ultrafast multislice helical CT scanning has replaced older techniques and allows the whole liver to be assessed in a few seconds. During these scans the patient is moved through the scanner gantry while both the X-ray source and X-ray detectors rotate around the patient. The X-ray detectors stream data to a computer, which uses complex mathematical algorithms to generate volumetric datasets based on the attenuation values of the different tissues. This dataset can be manipulated in a variety of ways in order to present the anatomy in the manner which best suits the clinical scenario.

CT scanning is particularly valuable in the diagnosis of focal liver lesions such as tumours, cysts or abscesses (Figure 24.13). It is also useful in the assessment of traumatic liver injury as haematoma, laceration and active bleeding may all be identified.

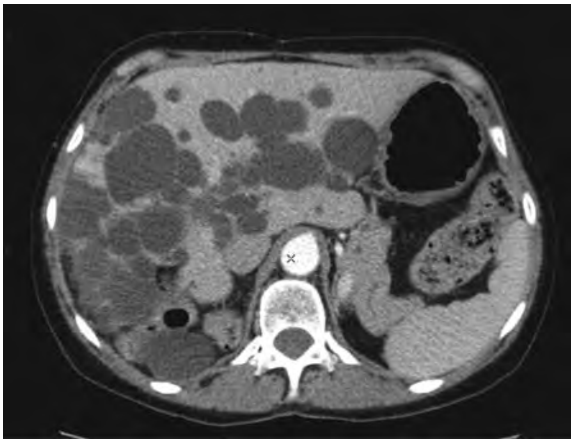


Figure 24.13 Axial CT scan image demonstrating multiple well-defined, round, uniformly fluid attenuation cysts in the liver of a patient with adult polycystic liver and kidney disease.

In modern practice, CT scanning is almost invariably assisted by the administration of intravenous iodinated contrast agents to 'enhance' both the parenchyma and blood vessels. Increased accuracy for detection of focal liver lesions is the principal benefit. In addition, focal liver lesions may also be characterized by observing the change in enhancement behaviour over time, relative to the surrounding liver parenchyma.

Specific timings of CT scan acquisition after administration of intravenous contrast may be employed to demonstrate suspected pathologies to best advantage. For example, lesions which characteristically receive a relatively abundant arterial blood supply relative to normal liver tissue [e.g. HCC, adenoma and neuroendocrine tumours (NETs)] may become conspicuously hyperenhanced during an 'arterial' phase approximately 25–35 seconds after administration of intravenous contrast. Such lesions may become inconspicuous during other phases of enhancement. On the other hand there are many liver lesions which are best demonstrated by virtue of their relatively poor blood supply compared with normal liver tissue. For instance, adenocarcinoma metastases from stomach, small bowel and colorectal primary tumours are most conspicuous during the 'portal venous' phase of enhancement approximately 65–75 seconds after administration of intravenous contrast.

These examples serve to illustrate the importance of tailoring the specific parameters of the CT scan to the specific circumstances of each patient. Effective communication between clinician and radiologist is essential in order to maximize the diagnostic utility for the patient.

MRI of the liver

MRI is an extremely useful tool in the assessment of both parenchymal liver disease and biliary disease, where it has superseded and replaced non-therapeutic endoscopic retrograde cholangiography.

The free protons in patients placed inside the bore of an MRI scanner align themselves with the static magnetic field of the scanner. These spinning protons are deliberately disrupted by pulses of radiofrequency energy which knock them out of alignment and out of synchronization. As the protons revert back to their original synchronization and alignment they themselves emit radiofrequency energy which is detected by receiver coils placed around the patient: these are analogous to aerials picking up conventional radio broadcast transmissions. The received radio signals are converted by computer algorithms into diagnostic images.

Different methods of applying radiofrequency pulses accompanied by different timings of 'listening' to the emitted radiofrequency energy are referred to as 'acquisition sequences'. Most liver imaging using MRI is produced using acquisition sequences which can be completed during a single breath-hold. One exception is free-breathing T_2 weighted cholangiography acquisition sequences which monitor the position of the diaphragm during the acquisition phase: the image reconstruction algorithm discards the data acquired while the diaphragm is moving and produces images only from data acquired while the bile ducts are stationary at end expiration.

Two common examples of image types produced by different acquisition sequences are known as T_1 and T_2 'weighted' images. T_1 weighted images reflect the fat content of tissues, among other features, and may be used to assess hepatic steatosis and examine focal liver lesions for the presence of fat which is characteristic for tumours of hepatocyte origin. T_1 weighted images may also be 'enhanced' by the administration of a variety of rare earth element-containing intravenous contrast agents which appear white on these images. T_2 weighted imaging produces images which reflect the amount of free water in the tissues, thus bile and other fluids appear white while denser tissues such as normal liver parenchyma appear darker grey. Magnetic resonance cholangiopancreatography (MRCP) images are produced using heavily T_2 weighted acquisition sequences. T_2 weighted images may also be 'enhanced' by the administration of a variety of iron-containing intravenous contrast agents, which appear dark on these images.

A variety of different intravenous contrast agents may be employed to enhance liver MR images. One class of contrast agents is based on the paramagnetic properties of rare earth elements such as gadolinium and manganese which affect the way protons realign themselves with the static magnetic field in the MRI scanner during some types of acquisition sequence. This type of MRI contrast agent modifies image appearances on T_1 weighted images. The rare earth contrast agents are further subclassified into conventional gadolinium chelate agents which enter the blood pool and subsequently become dispersed non-selectively throughout the extracellular space: these agents behave in a fashion similar to that of the iodinated intravenous contrast agents employed in CT scanning. During multiple phases of image acquisition these agents produce patterns of liver lesion enhancement and washout which are essentially the same as those observed during CT scanning (Figure 24.14). Another subclass of rare earth contrast agents are those which are selectively taken up by hepatocytes and subsequently excreted into bile. Immediately following intravenous administration these hepatocyte-specific contrast agents are located in the blood pool and influence tissue appearances in the same way as the conventional gadolinium chelate agents. However, their subsequent uptake by hepatocytes allows for more accurate characterization of focal liver lesions than can be achieved with conventional gadolinium chelate contrast agents. During a further late phase, when the hepatocyte-specific contrast agent is excreted into the bile, the bile becomes white on T_1 weighted images, thereby facilitating the identification of bile leaks (Figure 24.15) and assisting evaluation of the patency of biliary

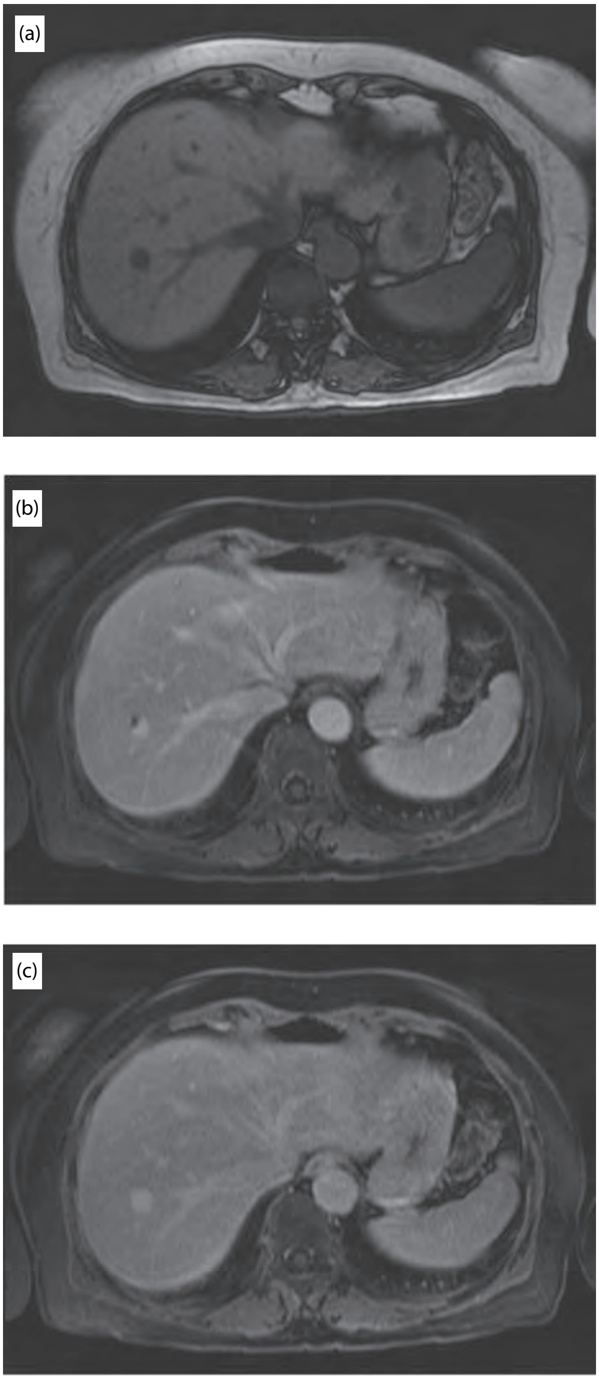


Figure 24.14 Axial MR images illustrate the behaviour of a conventional gadolinium-chelate intravenous contrast agent during dynamic imaging of a typical small cavernous haemangioma. The lesion appears as (a) uniformly low signal on the precontrast T_1 weighted image, then (b) fills in a globular fashion from the periphery before (c) becoming uniformly high signal several minutes after contrast administration.

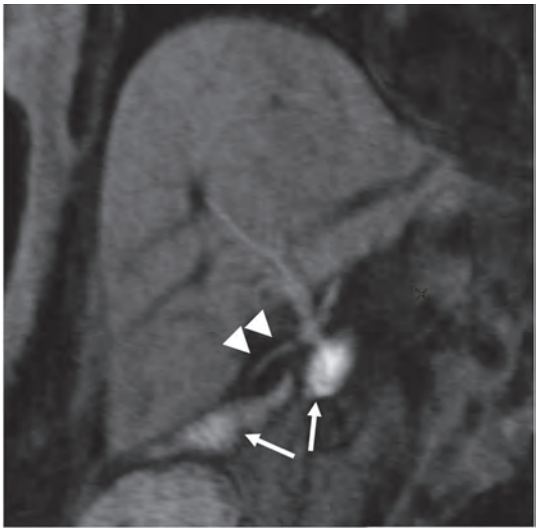


Figure 24.15 Coronal MR image illustrates the use of a biliary excreted hepatocyte-specific contrast agent to identify a bile leak at a biliary-enteric anastomosis. Contrast-enhanced bile traverses the anastomosis to enter the jejunal loop (long arrows) and also leaks laterally from the anastomosis and enters the lumen of an adjacent surgical drain (arrowheads).

enteric anastomoses. These two features are of great value to the liver surgeon.

The other class of contrast agent is based on the superparamagnetic properties of iron oxide which affect the way the spinning protons lose their synchronization during some types of acquisition sequence. This type of MRI contrast agent modifies images on so-called $T_2^\star$ weighted images. These contrast agents rely on uptake by cells of the reticuloendothelial system. In the liver normal liver parenchyma and some focal liver lesions contain Kupffer cells. On $T_2^\star$ weighted images the signal intensity from these tissues drops dramatically after administration of these reticuloendothelial contrast agents, rendering lesions which do not contain Kupffer cells conspicuous against a dark background. This behaviour facilitates both the detection of focal liver lesions which do not contain Kupffer cells (e.g. metastases) and also the characterization of focal liver lesions which otherwise display similar imaging characteristics on ultrasound, CT and MRI (e.g. hepatic adenoma, which does not contain Kupffer cells, and FNH lesions, which do contain Kupffer cells).

Recent advances in MRI of the liver include the application of diffusion-weighted imaging and MR spectroscopy. Diffusion-weighted imaging highlights the restricted potential for diffusion of water in some pathological tissue states and may aid identification and characterization of focal liver lesions. MR spectroscopy reveals varying levels of metabolites between normal and abnormal liver tissue and by means of this may assist characterization of focal liver lesions.

The following is a summary of contrast agents used in MRI of the liver (Figure 24.16):

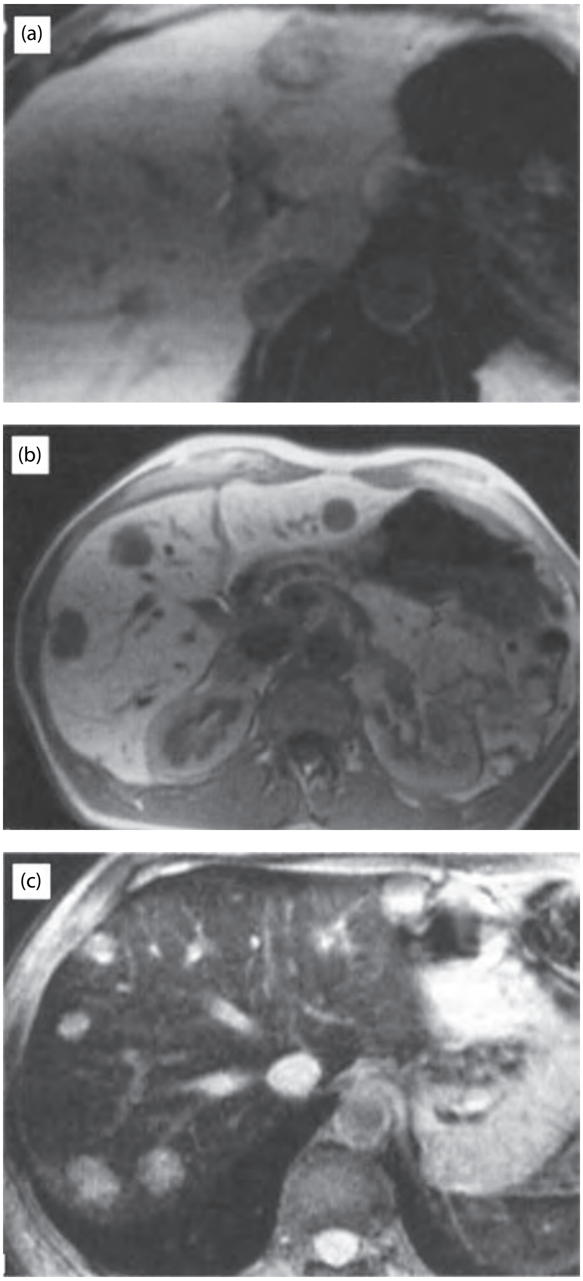


Figure 24.16 MRI scanning for liver metastases: (a) delayed T_1 weighted image following Gd-BOTA, (b) delayed T_1 weighted image following Mn-DPDF, (c) T_2 weighted image following superparamagnetic iron oxide – as the tumour deposits do not contain Kupffer cells, they appear as bright white lesions against the black (negatively enhanced) normal hepatic parenchyma.

- *Gadolinium (Gd) chelates with extracellular distribution (e.g. Gd-DTPA)*–following intravenous injection, these leave the intravascular compartment rapidly to equilibrate in the interstitial space. This limits their usefulness in the detection of focal hepatic lesions, but the technique is valuable in characterizing the nature of identified lesions by the different patterns of enhancement and washout, e.g. persistence of enhancement on delayed images in haemangiomas, strong initial transitory enhancement followed by delayed enhancement in FNH.
- *Hepatobiliary contrast agents (e.g. Gd-EOB-DTPA)* – these are hepatocyte specific and thus provide prolonged enhancement of normal liver parenchyma and lesions that contain hepatocytes. Thus HCC, FNH and the regenerative nodules of cirrhosis enhance with these agents, whereas secondary deposits, lymphomas and cholangiocarcinomas do not.
- Macrophage–monocyte–phagocyte system (MMPS) target agents – these consist of superparamagnetic iron oxide (SPIO) particles

coated with dextran that are cleared from the blood by the macrophage-monocyte system following intravenous injection. In the liver, they are taken up by the Kupffer cells abolishing the signal on heavily weighted T_2 imaging. As focal lesions (tumours in particular) do not contain Kupffer cells, they appear as bright white lesions against the black (negatively enhanced) normal hepatic parenchyma. This contrast imaging provides one of the best techniques currently available for the detection and precise location of hepatic secondary deposits particularly with three-dimensional reconstruction from the volumetric dataset.

- *Intravascular (blood pool) contrast agents* – ultra - small superparamagnetic dextran - coated particles (USPIO) are used as intravascular contrast agents in view of their long half-life in the blood (200 minutes). They are useful in the detection of hypervascular lesions (HCC, deposits from carcinoid tumours, haemangiomas, etc.) using dynamic MRI in much the same fashion as dual helical CT.

Ultrasound vs CT vs MRI scanning for liver imaging

Focal liver lesions

CT and MRI guarantee a comprehensive evaluation of the whole liver which may not be possible with ultrasound scanning depending on the body habitus of the patient. Ultrasound is also notoriously dependent on the skill of the operator, so reproducibility may be an issue in follow-up of challenging focal liver lesions. Nevertheless, in many situations, particularly for screening of high-risk patients and for follow-up of readily visualized focal liver lesions, the convenience and low cost of ultrasound may justify its use in preference to more costly MRI and CT which use ionizing radiation. CT and MRI may both be employed to detect and characterize focal liver lesions. Where focal liver lesion characterization proves challenging MRI has advantages over CT in that it allows for a greater range of tissue characteristics to be demonstrated, and hepatocyte-specific contrast agents may provide additional information which cannot be yielded by other means. Ultrasound using intravenous contrast agents may also assist the characterization of focal liver lesions. CT scanning retains a particularly valuable role in the evaluation of the profoundly unwell patient for whom the long duration and long breath-hold requirements of MRI scanning preclude practical application.

Diffuse liver disease

Diagnosis and evaluation of diffuse liver disease with non-invasive imaging techniques is a contentious area. There is no universally accepted technique, although some specialized methods involving ultrasound elastography are under evaluation. There are also MRI-based imaging techniques which may prove useful in this challenging diagnostic area.

Gallbladder and bile duct disease

Ultrasound remains the workhorse of suspected inflammatory gallbladder disease, and may be supplemented by CT evaluation if there is a clinical suspicion that sepsis has extended beyond the gallbladder. Ultrasound has limited utility in evaluation of the rest of the biliary tree in large part because it is only very rarely that the lower half of the common bile duct may be adequately visualized. For complex bile duct disease MRI using T_2 weighted acquisition sequences (MRCP) has become the technique of choice for high-quality non-invasive evaluation of suspected pathology of both the intra- and extrahepatic bile ducts. Coupled with anatomical information from T_1 weighted imaging the current ability of MRI to depict the extent of hilar cholangiocarcinoma allows the surgeon to determine the feasibility of resection in cases without resorting to trial of dissection. MRI in the delayed phase after administration of biliary excreted hepatocyte-specific contrast agents is the best non-invasive technique for detection and localization of bile leaks.

Conventional radionuclide scanning (scintigraphy)

Conventional radionuclide imaging produces images by detecting the radioactive decay of exogenously administered 'tracer' radioisotopes. These are usually tagged onto carrier compounds, thereby forming radiopharmaceutical agents, which direct the radioactive molecule to the target organs.

Advances in ultrasound, CT scanning and MRI, which produce images in far higher resolution than can be achieved with radionuclide imaging, have led to a reduction in the use of these techniques in the liver.

One specific radionuclide technique that retains a role is radiolabelled sulphur colloid scanning. Particles of sulphur colloid which have been labelled with the tracer technetium-99m are taken up by functioning cells of the reticuloendothelial system. In the liver these are the Kupffer cells, which are present in FNH lesions but not in hepatic adenomas. Radiolabelled sulphur colloid scanning may be useful to discriminate between these two focal liver lesions when the results of other imaging tests have been equivocal. Generalized impairment of hepatic tissue may be reflected in generally poor uptake of radiolabelled sulphur colloid when compared with other tissues containing large numbers of reticuloendothelial cells (spleen and bone marrow). In Budd–Chiari syndrome this generally poor uptake may spare the caudate lobe.

Another conventional radionuclide test, which retains a place in the diagnostic imaging toolbox, is functional biliary imaging with HIDA scanning. This test produces images based on the biliary excretion of intravenously administered ^{99m}Tc-labelled hepatobiliary iminodiacetic acid. In normal subjects the gallbladder is seen to accumulate the agent. If the gallbladder is not visualized within 4 hours of administration, then we can conclude that the cystic duct is obstructed. This finding may be useful in the assessment of patients whose symptoms point to a diagnosis of cholecystitis but in whom other imaging test findings are equivocal.

Positron emission tomography

Until fairly recently positron emission tomography (PET) has been used to investigate single organs, particularly the heart (myocardial perfusion, oxygen extraction, etc.) and the brain (metabolism, neurotransmitter function, etc.). With the advent of whole-body PET scanning, the technique is being introduced in oncology where it has potential and advantages over other imaging modalities:

- In principle, whole - body PET can identity a tumour and indicate the presence of spread anywhere in the body at an early stage (micrometastases).
- It can differentiate between benign and malignant tumours (metabolic activity of the lesion).
- It differentiates viable from non-viable tumour, which may be increasingly important for *in situ* ablation of tumours, e.g. liver, prostate.

PET scanning produces images by detecting the -radiation resulting from decay of positron-emitting radioisotopes. Positrons are short-lived positively charged electrons which generate a matter–antimatter annihilation interaction when they encounter electrons in the body. PET images are formed by detecting the -radiation emitted by these interactions. A unique feature is that each annihilation event generates two identical -rays which travel in diametrically opposed directions. PET scanners detect these pairs of -rays and use complex computer algorithms to produce a map of radioisotope distribution in the tissues.

The commonest positron-emitting isotope in current use is fluorine-18, which is usually administered in radiopharmaceutical form as ^{18}F-fludeoxyglucose. This is a glucose analogue which is taken up by actively metabolizing cells. The amount of uptake of this radiopharmaceutical in any tissue is proportional to the overall metabolic activity within the tissue. Thus lesions which are more or less metabolically active than surrounding normal tissues may be identified and differentiated. PET scanners are usually coupled with CT scanners to produce 'fused' images which allow the map of metabolic activity to be overlaid onto anatomical images produced by CT scanning.

The principal application of this technique in the liver itself is to aid discrimination between benign and malignant lesions. This discrimination is based on the assumption that malignant lesions are more metabolically active than benign lesions, although this may not always be the case. A further application is to detect residual viable tumour tissue at sites of resection or *in situ* ablation. However the commonest application of ^{18}F-fludeoxyglucose PET scanning in liver surgery practice is to detect previously unsuspected extrahepatic disease in patients with malignant liver lesions who are being considered for potentially curative hepatic resection or ablation therapies.

Although PET scanning of the liver is certainly a promising technique, further evidence and data are needed to fully establish its usefulness in clinical practice, its accuracy and specificity in different clinical scenarios.

Invasive imaging of the liver

In the past a variety of invasive interventional radiology techniques were employed in order to produce images of vascular structures in and around the liver and assess portal venous pressure. These have been almost entirely superseded by advances in the imaging techniques previously discussed. Modern interventional radiology applications in the liver are invariably performed with therapeutic intent (as is endoscopic retrograde cholangiography) and are guided by prior high-quality non-invasive imaging examinations. The commonest example is percutaneous transhepatic cholangiography performed to relieve obstructive biliary lesions by insertion of stents across obstructions and drainage catheters to facilitate resolution of biliary sepsis. Diagnostic brushings from obstructing biliary lesions may be taken at the same time and these may confirm the presence of malignancy on cytological examination. Hepatic arterial angiography may be performed to target arterial administration of therapeutic pharmaceutical agents in patients with HCC; a procedure usually referred to as transarterial chemoembolization (TACE).

Needle biopsy of the liver and of focal liver lesions

Despite advances in non-invasive imaging techniques imaging-guided biopsy still retains a valuable role in the evaluation of both diffuse liver disease and focal liver lesions. Please see the section General considerations (above) for further discussion of the place of liver and liver lesion biopsy in modern clinical practice.

The indications for biopsy in suspected or confirmed diffuse liver disease are where there is doubt as to the diagnosis, or where it is clinically important to determine the degree of advancement of the disease process. The indications for biopsy of focal liver lesions are where there is clinical concern that a potentially life-threatening pathology may be passing undiagnosed at a stage where there is still potential for cure, or where potential for cure has passed but tissue confirmation is required before embarking on palliative therapy. Occasionally biopsy is indicated when patients demand confirmation of diagnosis for peace of mind.

There is some debate about the appropriateness of biopsy in cases where a focal liver lesion may be a potentially life-threatening tumour but has not been characterized with absolute confidence using non-invasive imaging techniques. The concern arises because of the risk of causing tumour cell dissemination into the peritoneal cavity and chest or abdominal wall during the biopsy procedure itself. The perceived risk of such dissemination may be higher than the actual incidence of this phenomenon, but it is generally accepted practice that, for instance, a potentially resectable lesion which is likely to be a HCC should not be biopsied prior to attempting surgical resection.

Imaging-guided liver biopsy may be performed as a day case and carries a very low mortality rate (less than 0.1%). Evaluation of diffuse liver disease can only be performed on core biopsies obtained using Tru-Cut-type needles. Similarly for evaluation of focal liver lesions core biopsies are far preferable to fine-needle aspiration cytology samples.

Biopsy for diffuse liver disease may be performed without imaging guidance. However, ultrasound is so ubiquitous and so easy to use that it seems sensible to select a target site for biopsy using this imaging modality to ensure that, for instance, no lung tissue is traversed and that there are no unsuspected loops of bowel between the chest or abdominal wall and the liver.

Biopsy of focal liver lesions is most readily performed using ultrasound to guide the biopsy needle to the target tissue.

Histopathologists particularly appreciate biopsy cores which include the transition from normal to abnormal tissue and obtaining these requires practised technique. Highly skilled operators may even successfully biopsy lesions which are undetectable on ultrasound by targeting tissue using the relationship of the lesion to vascular structures as demonstrated on CT scanning or MRI.

All liver biopsies should be performed under conditions which minimize the likelihood of haemorrhagic complications: this includes the correction of any coagulopathy. If this cannot be achieved a transjugular biopsy may be performed to sample liver parenchyma in cases of diffuse liver disease: in this technique bleeding from the biopsy site is immaterial as the blood simply enters the systemic venous circulation. As well as haemorrhage other rare potential complications of liver biopsy include haemobilia, biliary peritonitis and hepatic arteriovenous fistula.

The most common complications of liver needle biopsy are:

- pleural effusion
- haemorrhage from the liver and thoracic wall
- intrahepatic haematoma
- hepatic arteriovenous fistulas
- haemobilia
- accidental puncture of the gallbladder and large bile ducts leading to bile peritonitis
- tumour cell implantation.

Laparoscopy with contact ultrasonography of the liver

Laparoscopy with contact ultrasound is an invaluable investigation that provides direct information on the state of the liver and the peritoneal cavity. It is particularly useful in the investigation of patients with jaundice, chronic liver disease, ascites of unknown origin and in the diagnosis and staging of both primary and secondary hepatic tumours. The advent of laparoscopic contact ultrasonography using high-resolution linear array probes has added considerably to the diagnostic yield of laparoscopy for the detection and staging of hepatic tumours. In addition to detecting secondary hepatic deposits in the liver and peritoneal cavity that are too small to be detected by CT scanning and MRI, laparoscopic contact ultrasonography permits the visualization of the hepatic vasculature (portal vein, hepatic artery, hepatic veins and retrohepatic vena cava) and examination of the entire biliary tract. This accurate assessment of the extent of involvement of the liver by tumour and relation to hepatic veins enables surgical planning of the resection necessary in the individual case.

Laparoscopy with contact ultrasonography and colour Doppler readily identifies haemangiomas. Visually guided biopsies or fine-needle aspiration cytology of both the hepatic parenchyma (in cirrhosis) and focal lesions (tumours) for histological confirmation carries a higher diagnostic yield than percutaneous liver biopsy, and should bleeding occur from the biopsy site, this can be controlled by compression or electrocoagulation.

In many centres it is routinely used to guide and monitor *in situ* ablation treatments like radiofrequency or microwave ablation (see below) and it is routinely employed during laparoscopic liver resection surgery (see below).

Clinical features of liver disease

Acute liver disease

Signs and symptoms

Patients may be asymptomatic or complain of non-specific symptoms, e.g. fatigue, malaise, anorexia, nausea, headache, myalgia, arthralgia and fever which commonly occur in hepatitis. Jaundice may develop in the later stages of illness.

There may be few signs apart from jaundice, best seen in the sclera, and hepatomegaly. Dark urine and pale stool are present in cholestatic disorders. Acute liver failure patients may develop features of hepatic encephalopathy with disease progression.

Chronic liver disease

Symptoms

This may be asymptomatic or patients may report non-specific symptoms, most commonly fatigue. Anorexia and nausea are common in all forms of hepatic and biliary disease and may be accompanied by weight loss and malnutrition. Jaundice is the most obvious symptom and in many instances will be accompanied by dark urine, pale stools and pruritus. Confusion, forgetfulness, poor concentration and personality change are central nervous symptoms of advanced hepatic cirrhosis [portosystemic encephalopathy (PSE)]. Fluid retention, ascites and dependent oedema, or haematemesis and melaena from gastrointestinal bleeding are manifestations of portal hypertension. Patients may complain of right hypochondrial pain due to hepatic distension or generalized abdominal discomfort due to ascites. Endocrine dysfunction may cause amenorrhoea, loss of libido or gynaecomastia.

Signs

General examination may reveal jaundice in the sclerae and skin. Pruritus may lead to scratch marks and may be present earlier than other manifestations of liver disease, especially in primary biliary cirrhosis. Xanthelasma and hyperpigmentation are also features very common in Primary biliary cirrhosis.

Chronic liver disease may show stigmata in the form of palmar erythema, finger clubbing, leuconychia, spider naevi, particularly over the upper half of the body, muscle wasting, loss of axillary/pubic hair and gynaecomastia. Central nervous system effects may be apparent with a liver flap, peripheral neuropathy and loss of consciousness ranging from confusion to full hepatic coma. In advanced disease with significant shunting, a sweet musky odour may be apparent on the breath: foetor hepaticus due to unmetabolized mercaptopurines; this is commonly associated with encephalopathy but is not pathognomonic and may occur in patients in clear consciousness.

On abdominal examination, there may be evidence of collateral veins in the abdominal wall, distension of the abdomen and eversion of the umbilicus from ascites, dependent oedema and testicular atrophy. On palpation, hepatomegaly should be searched for and a decision made whether this is real or apparent, whether it is diffuse, focal or multifocal. Liver tenderness may be elicited either abdominally or by palpation or through the rib cage by percussion. Auscultation may rarely elicit a friction rub over an expanding tumour or abscess. A systolic hepatic bruit is sometimes heard over a HCC or in patients suffering from alcoholic hepatitis. Splenomegaly may accompany hepatomegaly in patients with portal hypertension, it is rarely massive and frequently absent in patients with significant portal hypertension.

Parenchymal disorders of the liver

A full description of all liver disease is beyond the scope of this chapter. The management of parenchymal liver disease is predominantly medical, although some complications require surgical intervention. Both acute and chronic parenchymal liver disease can present with features suggestive of surgical pathology; failure to recognize this can have serious consequences and so these conditions should be considered in any differential diagnosis. Equally anyone suffering from a chronic liver disease can develop another condition requiring surgical intervention, which requires appropriate precautions. Many disorders require joint management between hepatologist, hepatic surgeon and an interventional radiologist.

Acute and chronic liver disease

Acute liver disease is a disease process affecting the liver for less than 6 months, e.g. hepatitis A, and implies that the insult will resolve. Clearly chronic liver disease must start and therefore have a period of time when its duration is of less than 6 months; for most chronic liver disease the start is subclinical for alcohol-related liver disease, and presentation is that of the chronic condition. Some conditions can present acutely and become chronic, e.g. hepatitis B. Most patients with chronic liver disease have no overt signs or symptoms of liver impairment and are considered to have 'compensated' disease; those with liver impairment evidenced by jaundice, ascites, encephalopathy are classed as 'decompensated'. The term 'acute-on-chronic' liver disease has recently been postulated to describe patients with a cirrhotic process, with evidence of liver failure associated with a 'systemic inflammatory response syndrome'. The usefulness of this concept is still debated but it does identify a group of patients with a very poor prognosis. This syndrome can be provoked after surgery in patients with chronic liver disease. Specific treatments that are of benefit to these patients over and above standard of care for liver failure have yet to be defined and trialled, but early aggressive support of the circulation and treatment of infection appear to be important.

Acute liver disease

Viral hepatitis

Acute viral hepatitis presents with nausea, right upper quadrant pain and usually jaundice, so should be included in many differential diagnoses. However, many cases of acute hepatitis are subclinical and will be discovered incidentally on testing of LFTs. It can be caused by hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis E virus (HEV) and by hepatitis C virus (HCV). A milder hepatitis can also be caused by cytomegalovirus, the Epstein–Barr virus and the protozoan Toxoplasma. In some cases of hepatitis, serological testing will be negative, no other aetiology will be evident, and it is assumed that this condition has a viral aetiology, as yet unidentified. It is further assumed that there is more than one agent and the illness is termed seronegative hepatitis. A small subset of these patients develops fulminant hepatic failure (FHF). The management of all these conditions is supportive, with observation for the tiny minority who will develop fulminant liver failure and will need invasive support and consideration of liver transplantation. In surgical practice, universal precautions should always be employed, particular care should be taken in operating on patients who are known to be infected with bloodborne hepatitides, but this should in no way prejudice the care of the patient. HAV and HEV are spread by the faecal–oral route. HAV is becoming much less common but continues to give rise to epidemics. Although the disease is usually benign, complications including prolonged cholestasis and macropapular skin reactions are common. A vaccine developed from killed virus A propagated in fibroblast culture is available. It appears to be particularly effective when combined with recombinant hepatitis B vaccine. HEV is predominant in Asia but is now endemic in the UK, but still more common in patients with a clear travel history to Asia. It has a generally benign course, except in pregnancy where there is a significant mortality. Neither HAV nor HEV infection leads to chronic infection or chronic liver damage.

HBV, which was first isolated in 1966, has infected in excess of 350 million individuals worldwide. It is a leading cause of chronic hepatitis, cirrhosis and HCC. The hepatitis B vaccination programme is altering the epidemiology of hepatitis B. Thus infection in healthcare workers and homosexuals (who have accepted the vaccination programme) has fallen. In those countries that have adopted universal vaccination programmes, huge reductions in HBV prevalence have been achieved and a decline in the complications of chronic infection has already been observed. It should be stressed that the infectivity of HBV is much greater (eight times) than that of HIV.

Chronic HBV infection is defined as the presence of HBsAg persistent more than 6 months after infection; spontaneous clearance is still possible at this stage but runs at between 2% and 4% per annum. The risk of chronic infection depends on the age of acquisition: with adult infection less than 10% become chronically infected, whereas infection in the under fives has a greater than 40% risk of chronic infection. The assessment of chronic HBV infection involves HBV serology, but also HBV DNA and an assessment of liver damage by either biopsy or non-invasive markers such as sonoelastography. The recognition of mutant HBV now means that viral load is more important than 'e-antigen' status in predicting prognosis and therefore the need for treatment. The specific therapy of chronic hepatitis B disease is now possible, but is complicated and rapidly evolving and should be undertaken in consultation with expert centres. Therapy has one of two different aims, either viral eradication with defined therapy or viral suppressive therapy long term to prevent complications of cirrhosis or HCC. Currently eradication therapy is based on -interferon, which inhibits the replication of HBV, and treatment can lead to remission of the disease with clearance of the hepatitis B e-antigen in up to 30% and hepatitis B surface antigen often follows. -Interferon is administered subcutaneously in a pegylated formulation weekly for 12 months. Factors associated with better response to -interferon include:

- HBeAg+
- a raised aminotransferase concentration >100 IU/mL
- low values for HBV DNA
- liver biopsy showing moderate to severe inflammatory activity
- age <65 years.

Viral suppressive therapy should only be contemplated in those patients who have significant risk of developing damage from their HBV infection, as it is likely to be lifelong and expensive. The aim of therapy is to depress viral load below the limit of detection, which not only reduces the risk of disease progression but also substantially reduces the risk of viral resistance. Therapy is dependent on nucleoside or nucleotide analogues; there is no synergy between these drug classes but there is a different resistance profile. Older antiviral drugs included the nucleoside analogue lamivudine (3-thiacytidine); although some patients remain on this drug, up to 40% will develop resistant virus within 2 years, with viral escape. Newer drugs with higher genetic barriers to resistance are preferred as front-line therapy. HBV antivirals have revolutionized liver transplantation for HBV with much improved survival rates.

Hepatitis C infection becomes chronic in 80% of cases; of these, 20% will progress to cirrhosis within 20 years. Worldwide 200 million people are chronically infected, and in many parts of the world it is now the leading cause of HCC and the commonest indication for liver transplantation. The commonest routes of transmission are intravenous drug misuse, unscreened blood products (pre-1991 in the UK) and unsterile medical practices. There are six genotypes of the virus with geographical variation: genotypes 1 and 3 are predominant in the UK, Europe and the USA. HCV infection is curable with combination therapy of pegylated -interferon and ribavirin in the majority of cases. Genotype 2 and 3 patients are easiest to treat, needing 6 months' therapy to achieve cure in 70% and 90% respectively. Genotype 1 patients, with a year of combination therapy, have cure rates of 40%; however, very shortly genotype-specific protease inhibitors will increase this to 70%. Following transplantation in those not cured pretransplant, nearly all patients become reinfected and 20% develop cirrhosis in the transplanted liver within 5 years.

Drug-induced liver injury

Since one of the main functions of the liver is to detoxify or to metabolize many pharmaceutical agents, it is perhaps not surprising that an overdose or abnormal response to the agent may lead to problems. It is not usually the drug but a metabolite that causes the liver cell damage, with overt jaundice and in some cases acute hepatic insufficiency (fulminant liver failure). There are many drugs that cause liver damage to varying extents. These may be classified as:

- directly hepatotoxic agents, e.g. carbon tetrachloride, tetracyclines, paracetamol, DDT and benzene derivatives
- drugs that interfe with bilirubin metabolism/transport:
  - haemolysis, e.g. para-aminosalicylic acid and phenacetin
  - impaired bilirubin excretion, e.g. methyl testosterone and norethandrone
  - interference with uptake and transport of bilirubin, e.g. rifampicin
  - interference with bilirubin conjugation, e.g. novobiocin
  - interference with bilirubin binding, e.g. salicylates and sulphonamides
- intrahepatic cholestasis, e.g. phenothiazine derivatives, chlorpromazine
- hepatitis-like disease, e.g. iproniazid, halothane, trichloroethylene
- hepatic fibrosis, e.g. methotrexate.

Damage is maximal in zone 3, where metabolizing enzymes are in the highest concentration and oxygen tension is the lowest. The histological picture may resemble acute hepatitis and, if so, has a poor prognosis. In other cases, light microscopy shows only scattered fatty change and no inflammation. A careful history of both prescribed drugs and self-medication is required for all jaundiced patients or those who have abnormal LFTs. From a clinical standpoint, drug-induced liver injury (DILI) has to be included in the differential diagnosis of most presentations of liver disease:

- cholestatic jaundice
- acute hepatitis
- chronic active hepatitis
- pseudoalcoholic liver disease
- hepatic sclerosis (due to vascular damage)
- hepatic neoplasia
- fulminant liver failure
- hepatic veno-occlusive disease
- vanishing bile duct syndrome.

DILI accounts for 40% of hospitalized cases of 'acute hepatitis'. Over 50% of cases of FHF are caused by drugs such as acetaminophen (paracetamol), halothane, phenytoin and methyl dopa. In surgical practice, hepatic injury induced by antimicrobial drugs is not uncommon, in particular co-amoxiclav (amoxicillin conjugated to clavulanate) can cause a hepatocellular or, more commonly, a cholestatic injury. The mechanism appears to be drug hypersensitivity to the clavulanic acid or a metabolite derived from it. Similar damage can be seen with flucloxacillin.

Halogenated anaesthetic agent-induced hepatitis

Although rare, halothane hepatitis is most likely to occur postoperatively and deserves mention. Although usually recoverable, severe hepatocellular damage progressing to fulminant liver failure is well documented. The injury appears to be due to hypersensitivity, and the impaired liver function is associated with pyrexia, skin rashes and eosinophilia. Importantly, the condition develops in patients after previous exposure to halothane. An antibody to a hapten metabolite of the anaesthetic agent (trifluoroacetate) is found in 30% of suspected cases of halothane hepatitis. Biochemically, there is elevation of the serum glutathione 5-transferase activity, and this appears to be a more specific and sensitive measure of anaesthetic-induced hepatic injury than the serum aminotransferases. Although there has been only one single case report implicating isoflurane and the hepatotoxic potential of this agent remains debatable, covalently bound antigens which are recognized by antibodies from patients with previous halothane hepatitis have been reported, indicating the distinct possibility of cross-sensitization with halothane or a common immune-mediated mechanism. Thus the practice of changing halogenated anaesthetics in patients requiring multiple anaesthetics does not guarantee a reduced risk.

Hepatotoxic herbal remedies

The range of DILI is not limited to prescription medication and can occur with over-the-counter medication and alternative remedies. Some herbal remedies are hepatotoxic and the spectrum of the disorders ranges from mild hepatitis to extensive necrosis, prolonged cholestasis, occlusion of the small hepatic veins (veno-occlusive disease), chronic hepatitis and cirrhosis. Chinese herbal teas (popular in the treatment of dermatitis) derived from the plants *Dictamnus dasycarpus* and *Paeonia* sp. have been confirmed to have hepatotoxic substances, and cases of severe jaundice and hepatitis have been reported after drinking these teas. Black cohosh used for menopausal symptoms and glucosamine used for joint pain have been associated with liver failure.

Ischaemic hepatitis

This condition, also known as 'shock liver', is the commonest cause of very elevated transaminases in hospital practice. It usually arises in older patients who have had an episode of hypotension. It is characterized by a rapid rise in transaminases into the 1000s within 24–36 hours of the insult and a rapid fall over the following days; often there is a secondary cholestatic phase lasting several weeks. In most patients these abnormalities are asymptomatic compared with the underlying illness, and the abnormalities resolve spontaneously. A small proportion of patients will become jaundiced; of these, some will develop a prolonged prothrombin time and even spontaneous hypoglycaemia. Management is supportive but the prognosis is poor and liver transplantation is usually contraindicated because of comorbidity that leads to shock liver.

Postoperative cholestasis

This is an old term coined more than 100 years ago with the advent of anaesthesia. It still has usefulness describing a syndrome of cholestasis seen in patients after major surgery and major sepsis and multiorgan failure in intensive treatment unit patients, but it is important to exclude drug reactions and ischaemic hepatitis. The aetiology is unclear but can be viewed as a toxic insult to the biliary canalicular system and epithelium from a combination of: bacterial products (from the gut or systemic sepsis), reduced perfusion and toxic bile. The reduced perfusion has been associated with the use of positive end-expiratory pressure in cardiothoracic procedures. The toxic bile may result from the downregulation of some canalicular transport systems in response to toxins. The clinical syndrome can vary from an asymptomatic biochemical cholestasis to prolonged jaundice or the development of a secondary sclerosing cholangitis.

Chronic liver disease

As described above, chronic liver disease has a duration of greater than 6 months. To some extent, depending on aetiology, it does not cause many symptoms or clinical complications until sufficient loss of liver parenchyma and increase in fibrosis has led to the development of cirrhosis. This and its aetiologies are described below; of course, viral hepatitis, both HBV and especially HCV, can cause chronic disease as described above.

Cirrhosis of the liver

Liver cirrhosis is the end result of hepatocyte death and regeneration with fibrosis. It must be distinguished from hepatic fibrosis, which can occur in the portal regions from chronic bile duct obstruction or congenitally, or around the central veins in chronic cardiac failure. Confluent necrosis of zones 1 and 3 leads to fibrotic bridges, and the regeneration of surviving hepatocytes results in a further distortion of hepatic architecture. There are multiple causes of cirrhosis. In the Western world, alcohol, hepatitis C and non-alcoholic steatohepatitis (NASH) would be the commonest causes. Irrespective of the underlying aetiology of the cirrhosis, the clinical syndrome and complications are very similar or identical, i.e. liver failure and portal hypertension.

Not uncommonly, the condition will be unsuspected and comes to light because of a routine estimation of LFTs or found incidentally at laparotomy. Clinical suspicion is aroused by finding stigmata of chronic liver disease such as palmar erythema, spider naevi, or otherwise unexplained peripheral oedema. Alternatively, the patient may present in a later stage of disease with muscle wasting and ascites, with gastrointestinal haemorrhage from varices, jaundice or hepatic encephalopathy.

In most patients, liver cirrhosis has a poor prognosis. However, if the underlying cause can be removed, return to normal function is possible with a huge improvement in prognosis, abstinence from alcohol being a typical example. Other diagnoses are less likely to regress, but in some the prognosis may be improved with appropriate management. Thus in the patient with chronic liver disease who presents with a condition requiring surgical intervention, not only is it important to be sure that the patient requires the intervention, but also is this the optimum time? Can the patient's condition be improved. If it can then the correct decision will often be to delay surgery.

Non-alcoholic fatty liver disease

NAFLD is the most common form of liver disease in the Western world. It is considered part of the metabolic syndrome, and as such is closely associated with type 2 diabetes mellitus, obesity, insulin resistance, hypertension and dyslipidaemia. The definition of non-alcoholic varies and in clinical practice a common-sense approach should be taken as to the main driver of liver damage: in general those drinking around 21–28 units of alcohol per week should be regarded as being within NAFLD.

NAFLD is an umbrella term encompassing a histological spectrum of disease ranging from simple steatosis to NASH. NASH can progress to the development and propagation of fibrosis leading ultimately to cirrhosis and HCC.

The pathology of NAFLD is characterized by excess fat within the liver. The excess fat is predominantly macrovesicular fat and usually seen in a perivenular (zone 3) pattern. Beyond simple fatty change, liver injury in NAFLD includes evidence of hepatocellular injury, most commonly seen as hepatocyte ballooning. Inflammatory changes and the development of fibrosis are also key histological changes seen in progressing NAFLD (or steatohepatitis).

To aid histological assessment, numerous scoring systems have been proposed. The most commonly used and widely accepted is the NAFLD activity score (NAS). The NAS is a semiquantitative assessment of the three cardinal histological features of NAFLD – steatosis, lobular inflammation and hepatocyte ballooning. Both steatosis and lobular inflammation are graded 0–3, with ballooning graded 0–2. The unweighted sum of these features constitutes the NAS, with score 5 diagnostic of NASH, score 3 considered borderline NASH and scores 2 labelled ‘not NASH’. Although the NAS system is not widely used in routine clinical practice, it is an essential component of many research studies in the NAFLD arena.

Clinical signs and symptoms are uncommon in NAFLD. When seen, the most common complaints are of vague right upper quadrant discomfort and fatigue. Most commonly, mildly abnormal elevations of aminotransferases are found incidentally in patients with some or all of the components of the metabolic syndrome. In patients undergoing obesity surgery, the prevalence of simple steatosis was 91% and 37% for NASH. NALFD is also more commonly found in patients with impaired glucose tolerance and overt type 2 diabetes mellitus, with up to 62% of newly diagnosed type 2 diabetes mellitus patients having coexisting NAFLD. As well as associations with the more commonly recognized features of the metabolic syndrome, NAFLD is also associated with other disease states including polycystic ovarian syndrome, hypothyroidism and colorectal adenomatous polyps.

There is currently no specific drug treatment; weight loss with an increase in physical activity is effective if it can be achieved. The prognosis is generally good for patients with steatosis, although they appear to have excess cardiovascular mortality. For those with NASH the prognosis is poorer with many progressing to cirrhosis within 10 years of diagnosis and those that do having an increased risk of developing HCC.

Alcohol-related liver disease

Alcoholic liver disease (ALD) is a major problem, especially in the affluent Western countries. The spectrum of liver damage varies from fatty infiltration to alcoholic hepatitis, hepatic fibrosis and cirrhosis. The last one is the commonest cirrhotic liver disease in some Western countries. The increasing per capita alcohol consumption is not just due to those who drink to excess drinking more but also to an increase in the median level of consumption; as a consequence, there has been a dramatic increase in the amount of alcohol-related liver disease. With the increase in the average amount drunk, many people now drink hazardously without signs of alcohol dependency but can still develop significant liver disease. However, it is important to note that only about 20% of those drinking at hazardous levels develop overt liver disease.

Despite intensive research, the exact mechanism responsible for the hepatic damage remains unknown; genetic factors must be important but so far polymorphisms of key genes have not been shown to confer susceptibility to ALD. Dietary deficiencies are not considered a major causal influence, but are highly prevalent as a secondary phenomenon and should be sought and treated. However, oxidative stress is a central mechanism of damage and a failure to adapt to this stress will predispose to alcohol-related liver disease. While the mechanism of activation of the stellate cells in ALD is the subject of much research, once activated they are the central players in the ongoing fibrosis of the liver in a similar manner to other liver diseases. Serum endotoxin levels are known to be increased in patients with ALD and stimulation of the phagocytes by endotoxin with excess production of tumour necrosis factor- has been suggested as a possible mechanism for hepatocyte damage. Acetaldehyde derived by oxidation from ethanol in the liver is involved in the increased synthesis of type I collagen in the liver, hence the fibrosis. Another factor which has been implicated in the fibrosis is the cytokine transforming growth factor (TGF)- , which is produced in excessive amounts by the Kupffer cells in response to alcohol. The excess TGF- stimulates the transformation of stellate into active fibroblasts. It is unclear yet which subtype of TGF- is implicated and although TGF- 1 has been widely reported, it is currently thought that more than one subtype is involved.

The presentations of alcohol-related liver disease can be divided into three: fatty liver disease, alcoholic hepatitis and cirrhosis, either compensated or decompensated. Alcohol-related fatty liver disease is now more commonly know as alcoholic steatohepatitis; it usually presents with incidental discovery of abnormal LFTs. Occasionally, there may be right upper quadrant discomfort and sometimes hepatomegaly, which may be tender due to fatty infiltration of the parenchyma. At this stage liver biopsy is not usually needed unless after screening for other liver diseases an alcohol history raises diagnostic uncertainty. If performed, it shows the hepatocytes to be distended with fat and sometimes there is an inflammatory infiltrate, early fibrosis or hyaline deposits (Mallory’s hyaline).

Alcoholic hepatitis is an acute severe condition that develops on the background of chronic alcohol use, either in a liver with normal architecture or more usually cirrhosis. The clinical features are of jaundice, often with ascites or encephalopathy, pyrexia and a bounding vasodilated circulation. Laboratory investigations reveal leucocytosis, raised inflammatory markers but only a mild elevation of the transaminases. The syndrome is defined by its poor short-term prognosis with untreated mortality of over 50%. The diagnosis can reliably be made on clinical grounds; some authorities advocate liver biopsy to confirm the diagnosis. The liver biopsy in alcoholic hepatitis shows a cellular infiltrate, liver cell necrosis, cholestasis, variable degree of fatty change and sometimes Mallory’s hyaline. These classical features can be seen in patients who are entirely well, adding to the controversy of biopsy. Treatment of alcoholic hepatitis includes cessation of alcohol and most advocate the use of corticosteroids guided by the Maddrey's discriminant function or the Glasgow alcoholic hepatitis score, which both identify high-mortality groups. Ongoing trials are comparing the use of pentoxifylline in combination with steroids or as monotherapy. The differential diagnosis is of end-stage liver disease with no inflammatory components.

The final stage is the development of cirrhosis, which is usually of the micronodular variety, but this is no longer a useful concept. The development of cirrhosis occurs some time in advance of clinical consequences, so may not be obvious until a hepatic insult such as surgery causes it to decompensate and manifest as liver failure with jaundice, ascites or encephalopathy. The most important factor in the treatment of ALD is abstinence from alcohol, and considerable improvement is seen both in LFTs and on repeat biopsy. Enteral nutrition imparts significant benefit in patients with alcoholic hepatitis and may improve survival, although this is debatable. The evidence for specific nutritional aids or supplements is poor and the control arm of the studies, if any, often represents inadequate standard hospital nutrition, so ensuring adequate dietary intake is clearly better than starvation for these patients. Transplantation is now performed for advanced alcohol-related chronic liver disease. Minimum periods of 6 months' abstinence before transplantation have previously been applied; however, these are currently being challenged as many patients with most to gain from transplantation do not survive that long. Also, due to the ill health of the patients, the abstinence is less meaningful as the patients are hospitalized for much of the duration. The results to date have been better than predicted and only about 15% of patients have resumed drinking after transplantation.

There are no specific laboratory markers of ALD, although the following are suggestive: raised levels of γ-GT, presence of carbohydrate-depleted (sialic acid depleted) transferrin and a specific isoenzyme of alanine aminotransferase (F-AAT). However, none is specific enough to rely on in an individual case, but they can raise the index of suspicion of an alcohol-related aetiology.

Cholestatic liver diseases

Cholestasis, impaired bile secretion/excretion, with the development of conjugated hyperbilirubinaemia and raised bile ductular enzymes can be caused by obstructive lesions of the biliary tract such as stones and tumours. In addition, the same pattern can be seen without dilated bile ducts and is described as intrahepatic cholestatsis. The causes of this include drugs, alcohol, viral hepatitis, primary biliary cirrhosis and primary or secondary sclerosing cholangitis. The term 'vanishing bile duct syndromes' is sometimes applied collectively to these disorders; it is a histopathological observation of ductopenia, or a paucity of bile ducts, and is not a specific disease. It can be observed in the conditions listed above and also hepatic sarcoidosis, graft-vs-host disease and hepatic transplant rejection, but is only present in a minority of cases.

Primary biliary cirrhosis

This is a disease of unknown aetiology in which the intrahepatic bile ducts are progressively destroyed by an immunological process. It runs a variable course but ultimately ends in primary biliary cirrhosis (as distinct from secondary biliary cirrhosis due to chronic obstruction of the extrahepatic biliary tract, e.g. strictures). Circulating antibodies against mitochondrial constituents (antimitochondrial antibodies) are found in all patients. These antibodies are non-organ and non-species specific and their relationship to the aetiology of the disease remains speculative but they are useful in the diagnosis of the condition. The most widely held immunological hypothesis for the development of the disease is aberrant expression of class II histocompatibility antigens on the epithelium of the bile ducts which induces a T-cell-related progressive immune destruction. There is some evidence that the mitochondrial antigens responsible for the production of antimitochondrial antibodies (e.g. E_2 antigen, which is also present in Gram-negative bacteria) may cross-react with the bile duct antigens.

The disease is often asymptomatic for long periods. Some cases are discovered accidentally because of abnormal LFTs obtained before blood donation. There is a female preponderance and the mean age at symptomatic presentation is 40 years. The symptoms include itching, weight loss, malaise and icterus. The liver becomes enlarged and, with progression of the disease, portal hypertension with splenomegaly develops. There is deposition of cholesterol in the tissues especially around the orbits and on the extensor surface of the large joints. In advanced disease, intrapulmonary shunting is associated with finger clubbing. The malabsorption secondary to the diminished bile salt pool leads to deficiency of fat-soluble vitamins (A, D, E, K) with the development of osteoporosis. In addition to biochemical features of cholestasis and presence of antimitochondrial antibodies, the serum IgM is elevated. Smooth muscle antibodies are also present in many patients.

There is no effective medical therapy but symptomatic relief is obtained by a variety of drugs. Cholestyramine is used for itching and ursodeoxycholate improves symptoms and the LFTs by replacing the toxic hydrophobic bile acids. Prednisolone seems to reduce fatigue and itching and may improve liver function. Monthly injections of fat-soluble vitamins are administered to counteract the deficiencies caused by the malabsorption. However, the disease process is not influenced by medical treatment and the only effective therapy that imparts a long-term cure is hepatic transplantation. Primary biliary cirrhosis is the second most frequent indication for liver transplantation. Nowadays, this operation is undertaken before the development of end-stage disease and the onset of hyponatraemia or significant bone disease. It is considered in patients whose quality of life has deteriorated or in whom the bilirubin exceeds 100 mol/L, and in those who develop portal hypertension.

Primary sclerosing cholangitis

Primary sclerosing cholangitis (PSC) is a chronic inflammatory large duct cholangiopathy, leading to fibrotic strictures of the intra- and extrahepatic bile ducts, of unknown aetiology. It is rare in the UK, but the leading indication for transplantation in Scandinavia. The clinical course of PSC varies considerably between patients, from decades of asymptomatic abnormalities in LFTs to frequent recurrent bouts of cholangitis progressing to liver failure requiring transplantation within a decade. Median survival from diagnosis to death or transplantation is 12 years, but the variability of the natural history makes prognostication difficult for the individual patient. The disease also carries an increased risk of cholangiocarcinoma occurring in 10–15% of cases. The diagnosis of this malignant stricture among the strictures in the biliary tree due to PSC is a significant challenge. The disease is commonly associated with inflammatory bowel disease, with up to 60% of PSC patients having inflammatory bowel disease if this diagnosis is searched for. The only curative treatment is liver transplantation. Ursodeoxycholic acid therapy is used on the basis that this is a cholestatic disease, but this is not recommended in guidelines and lacks any evidence base.

Autoimmune hepatitis

Autoimmune hepatitis is a disease characterized by persisting liver inflammation with piecemeal necrosis and an infiltrate of predominantly plasma cells. Classically, it presents in young women with a second peak in late middle age with a more even gender distribution. The disease is associated with autoantibodies; the antismooth muscle antibody is positive in 60% with the disease and the antinuclear factor is positive in 40%; both can be positive in other liver diseases at low titre. The liver, kidney and microsomal antibody are associated with a subtype of this disease that affects children. Presentation is often with incidental findings of abnormal LFTs and, on occasion, with jaundice and rarely variceal bleeding. The natural history without treatment is of progression to cirrhosis within 3–5 years; this may represent a historically more narrowly diagnosed disease. Treatment is with steroids and often a second-line immune modulator such as azathioprine (as a steroid-sparing agent); this therapy usually produces a rapid recovery and normalizes with at least 15-year survival.

Congenital and genetic liver diseases

Haemochromatosis

Congenital and genetic liver diseases are listed in Box 24.1.

Both genetic haemochromatosis and neonatal haemochromatosis represent a disorder of iron metabolism that is inherited as an autosomal recessive trait. They are thus distinct from secondary (acquired) haemochromatosis, which develops in patients with polycythaemia and those requiring multiple repeated blood transfusions (chronic haemolytic anaemias).

BOX 24.1 Congenital and genetic hepatic disorders

- Genetic haemochromatosis: defect of iron metabolism
- Neonatal haemochromatosis: defect of iron metabolism
- Wilson disease: defect of copper metabolism
- Cystic fibrosis
- α_1-Antitrypsin deficiency: sequestration of α_1-antitrypsin granules in endoplasmic reticulum
- Hereditary tyrosinaemia type 1: absence/deficiency of fumarylacetoacetate hydrolase
- Crigler-Najjar syndrome: absence/deficiency of uridine diphosphate-glucuronosyl transferase

Genetic haemochromatosis is common. The gene carrier rate in the North European population is 10% and the disorder may affect up to 1 in 300 heterozygotes. The genetic abnormality arises in the *HFE* gene and most commonly consists of a single point mutation, causing a single amino acid substitution from cysteine to tyrosine (C282Y). Another mutation causing an H63D substitution has a much weaker phenotype and usually only causes disease when compounded with C282Y. Together these account for 90% of phenotype haemochromatosis; however, it is important to note that the genotype has poor penetrance so only a minority develop phenotypic disease. The defect leads to an enhanced transport of dietary iron through the intestinal cell, and this is caused by failure of the normal mechanism of downregulation of the villous enterocyte transferrin receptors in response to excess iron stores. The discovery of the candidate gene has led to a simple effective screening method for the disorder.

If unrecognized, genetic haemochromatosis results in progressive iron deposition in the liver, heart, pancreas, joints and endocrine glands with sparing of the spleen, lymph nodes and bone marrow. The hepatic accumulation of iron leads to cirrhosis with an increased risk of HCC, which is one of the commonest causes of death in untreated patients. The classical clinical picture of the full-blown disease is that of a diabetic patient (75%) in whom there is a dusky brown pigmentation of the skin, buccal mucosa and conjunctiva (bronze diabetes) and cirrhosis. About half the patients have a polyarthropathy starting in the small joints, and many exhibit other endocrine dysfunctions including hypopituitarism and hypogonadism. The liver biopsy shows excessive iron deposition in the hepatocytes and the Kupffer cells with fibrosis or macronodular cirrhosis. The deposited iron is thought to induce hepatic injury by peroxidation of the intracellular phospholipid membranes. In addition, these patients are subject to cardiomyopathy.

Early diagnosis is essential to prevent the development of multiorgan disease. Characteristically, the serum ferritin is elevated and the transferrin saturation exceeds 55%, in the presence of which a liver biopsy is essential for establishing the diagnosis. Treatment of the disease entails phlebotomy and reduced dietary intake of iron with long-term follow-up of all patients.

Wilson disease

This is a rare autosomal recessive inherited disorder of copper metabolism leading to copper overload of various organs: liver, cornea, kidneys and the central nervous system (basal ganglia). The gene for Wilson disease (*ATP7B*) is located on the long arm of chromosome 13. Wilson disease is characterized by a low serum caeruloplasmin level. The basic cellular defect underlying the disease is unknown but there is a markedly decreased biliary excretion of copper, which is largely responsible for the progressive copper overload. The hepatic injury results in fibrosis and cirrhosis at a relatively young age and portal hypertension is common. Neurological features are more common in adult patients. The classical clinical feature is that of the Kayser–Fleischer ring of pigment in the cornea. Aminoaciduria and phosphaturia are the result of renal damage following copper deposition in the proximal tubules. Copper studies (serum copper, serum caeruloplasmin and urinary copper levels) should be undertaken in all young patients presenting with chronic liver disease, and slit-lamp examination of the cornea is also carried out. Although there may be difficulty in distinguishing Wilson disease from other chronic forms of liver disorders, as these are often accompanied by increased hepatic copper, the finding of a low plasma caeruloplasmin resolves this diagnostic difficulty.

The treatment of Wilson disease consists of chelation therapy with penicillamine or tientine together with avoidance of food high in copper content. Zinc supplementation is useful because it protects to some extent against the copper-induced hepatocyte injury and reduces the absorption of copper by the enterocytes. Ammonium tetrathiomolybdate is particularly effective for patients with neurological or psychiatric manifestations of the disease.

Cystic fibrosis

As more patients with cystic fibrosis are surviving to adulthood, the incidence of hepatic disease associated with this congenital disorder is increasing. Some 25% of patients with cystic fibrosis have clinical and biochemical evidence of liver disease. Many of these patients have intrahepatic and extrahepatic biliary strictures. Improvement in LFTs may be obtained by therapy with ursodeoxycholic acid, which increases the bile flow and alters the composition of bile such that the proportion of the less toxic hydrophilic bile acids in the bile acid pool is enhanced.

α1-Antitrypsin deficiency

This inherited disorder affects Caucasians of European descent. The deficiency of the α_1-antitrypsin (which acts as an antiprotease) results in pulmonary damage due to lack of inhibition of neutrophil elastase and to liver disease. The latter appears to be secondary to the sequestration of α_1-antitrypsin granules in the endoplasmic reticulum of the hepatocyte. There is no effective medical therapy and progressive disease is treated with hepatic transplantation.

Fibropolycystic diseases

Fibropolycystic liver disease encompasses a spectrum of related lesions of the liver and biliary tract that are caused by abnormal embryological development of the ductal plates. These lesions (congenital hepatic fibrosis, biliary hamartomas, autosomal dominant polycystic kidney disease, autosomal dominant polycystic liver disease Caroli disease, choledochal cysts) can be clinically silent or can cause signs and symptoms such as cholangitis, portal hypertension, gastrointestinal bleeding, infections and space-occupying masses.

Congenital hepatic fibrosis is a disorder that is inherited as an autosomal recessive trait, the importance of which lies in the development of portal hypertension with variceal bleeding in afflicted individuals. In congenital hepatic fibrosis, there is enlargement of the portal spaces by pronounced fibrosis and bile ductule proliferation. The ductules are dilated (ectatic) to varying degrees but still communicate with the main intrahepatic biliary tree. It is generally thought that the bile duct proliferation is the primary abnormality. Some of the bile ductules may become so dilated as to form communicating microcysts. The disease though diffuse is patchy with areas of unaffected liver, and essentially the overall architecture of the liver remains normal, although the portal fibrosis results in a block to the sinusoidal circulation within the liver and hence the development of portal hypertension.

Both sexes are affected and the disease is uncommon with a prevalence of 1:100 000. In most patients the condition is first recognized by the onset of severe gastrointestinal bleeding, which on endoscopy is found to be due to ruptured oesophageal or gastric varices usually between the age of 5 and 20 years. Less commonly, the condition presents with abdominal discomfort caused by an enlarged spleen or because of hypersplenism. In a small minority of patients, the presentation is with bacterial cholangitis. The clinical findings include hepatosplenomegaly with normal LFTs except for a slight elevation of the alkaline phosphatase or γ-GT in some patients. The liver is hyperechoic on ultrasound examination. The diagnosis is established by liver biopsy.

In 50% of cases congenital hepatic fibrosis is associated with Caroli syndrome (ectatic dilatation of the segmental bile ducts causing recurrent bacterial cholangitis). Other associated malformations that may occur include ectatic renal collecting tubules (similar to but distinct from medullary sponge kidney), duplication of the portal vein, cystic dysplasia of the pancreas, pulmonary emphysema, intestinal lymphangiectasia, cerebellar haemangioma, cleft palate and aneurysms of the renal and hepatic arteries.

The clinical course of congenital hepatic fibrosis is that of repeated episodes of variceal haemorrhage and as the liver function remains normal in these patients. A programme of banding therapy to achieve variceal eradication is usually effective at preventing bleeding. Placement of a transjugular intrahepatic shunt, especially a covered stent with improved long-term patency, is a good treatment option and may become the preferred long-term treatment for these patients. With their well-preserved liver function, surgical treatment by selective decompression by Warren shunt after the acute episode has been controlled by endoscopic banding is an option for selected patients.

Gilbert syndrome

Gilbert syndrome is the commonest cause of an isolated hyperbilirubinaemia. It is an unconjugated hyperbilirubinaemia and due to a mutation in the TATA box of the gene for bilirubin uridine diphosphate glucuronosyl transferase. This slows production of conjugated bilirubin leading to mild elevations of bilirubin that are not clinically apparent unless the patient is stressed or dieting. The condition affects 5–10% of the population. It is asymptomatic and is usually an incidental finding with the rest of the LFTs being normal. The diagnosis can be confirmed by genetic testing. In the acute setting it can cause diagnostic confusion, implying any liver problem is of more severity, especially when the acute pathology causes abnormalities in the LFTs.

Hereditary tyrosinaemia type I

This metabolic disorder is characterized by a deficiency of the enzyme fumarylacetoacetate hydrolase (FAH), which is the last enzyme involved in the degradation pathway of tyrosine.

The accumulated products of the amino acid together with various catalytic proteins result in renal, hepatic and neurological involvement. There are two clinical forms of this disorder. The acute type (total absence of FAH) is dominated by failure to thrive and progressive liver failure with death in infancy. The chronic variety (reduced levels of FAH) is characterized by renal tubular dysfunction, rickets, progressive liver disease and the development of HCC. The best treatment of this condition is hepatic transplantation, which abolishes the risk of HCC.

Crigler–Najjar syndrome

This congenital disorder which results in hyperbilirubinaemia is covered in Chapter 25.

Portal hypertension and complications

Portal hypertension arises as the result of obstruction to portal venous outflow; this may result from extrahepatic compression or thrombosis of the portal, mesenteric or splenic veins, from compression of portal venous radicles within the liver from a wide variety of liver diseases or from obstruction to the outflow from the liver (Box 24.2).

Rarely, anomalous arterioportal fistulas result in a massive rise in portal venous flow and pressure (see below). The degree of hypertension depends not only on obstruction to outflow but also on the volume of inflow; for obstructive disease such as portal vein thrombosis the increase in inflow is minimal so the portal hypertension is mild, whereas for cirrhosis there is a massively increased splanchnic blood flow, so even with incomplete obstruction there is significant portal hypertension.

The portal pressure is normally estimated in postsinusoidal cirrhosis by performing a wedged hepatic vein pressure and subtracting the inferior vena cava pressure, giving the wedged hepatic venous portal pressure gradient (PPG). A PPG <7 mmHg is functionally normal, and at 7 mmHg ascites and varices can occur, but varices will only bleed at PPG >11 mmHg. The measurement of PPG is a useful clinical tool but is not widely performed outside major centres. In sinusoidal and presinusoidal portal hypertension the PPG is normal and portal pressure must be measured by direct methods, such as splenic puncture, which these days are rarely done. Most commonly, portal hypertension is post sinusoidal and results from cirrhosis of the liver. Although a precise diagnosis may not be relevant to the immediate management of a patient with variceal bleeding, it may ultimately indicate prognosis and change the choice of subsequent treatment.

BOX 24.2 Pathogenesis of portal hypertension

- Increased blood flow into portal venous system (no obstruction)
  - Hepatic and splenic arterioportal fistulas (rare)
- Extrahepatic outflow obstruction
  - Hepatic vein thrombosis; Budd–Chiari syndrome, veno-occlusive disease; tricuspid incompetence, right heart failure
- Extrahepatic inflow obstruction
  - Congenital malformation of portal vein
  - Portal vein thrombosis
  - Splenic vein thrombosis (sectorial portal hypertension)
  - Portal vein compression, e.g. nodes
- Intrahepatic obstruction
  - Presinusoidal: periportal fibrosis and schistosomiasis
  - Postsinusoidal: cirrhosis (alcoholic, nutritional, postnecrotic, biliary), veno-occlusive diseases, haemochromatosis, Wilson disease, congenital hepatic fibrosis

Cirrhotic portal hypertension is the result of a combination of increased portal venous resistance and an increased splanchnic blood flow. Increased vascular resistance is both a mechanical consequence of liver architecture distortion and a dynamic process involving active contraction of myofibroblasts, activated stellate cells and venous smooth muscle cells. Vascular resistance may be modified by endogenous factors or pharmacological agents. Endothelin, -adrenergic stimulus and angiotensin II increase hepatic vascular resistance whereas nitric oxide, prostacyclin and vasodilating medications (e.g. nitrates, calcium channel blockers) reduce resistance. Splanchnic arteriolar vasodilatation is secondary to elevated circulating levels of vasodilator substances (including glucagon, nitric oxide) and decreased sensitivity of the splanchnic vasculature to endogenous vasoconstrictors. Patients have a hyperdynamic circulation with increased cardiac output, hypotension and hypervolaemia from the resulting sodium retention (discussed further under the section Ascites). Portal hypertension sometimes develops in patients with liver disease prior to the onset of cirrhosis. This has been attributed to intrahepatic arterioportal anastomoses, hyperdynamic circulation and the accumulation of vasoactive humoral factors which alter the resistance to flow.

About 25% of patients will have an extrahepatic cause for portal hypertension, usually a portal vein thrombosis. A proportion of these patients will also have underlying liver disease or hypercoagulability disorders such as polycythaemia. Chronic pancreaticobiliary disease or pancreatic neoplasms may be precipitating factors for portal or splenic vein thrombosis, but only rarely does neonatal umbilical sepsis seem to be an aetiological factor. Extrahepatic outflow block may result from thrombosis or occlusion of the hepatic veins (Budd–Chiari syndrome). Aetiological factors may be protein C deficiency, the contraceptive pill and ingested toxins that include *Senecio* or bush tea poisoning (veno-occlusive disease). Other patients have congenital diaphragms in the suprahepatic vena cava or chronic congestive right heart failure. Such patients rarely present with bleeding varices but suffer with intractable ascites, painful hepatomegaly and rapidly deteriorating liver function. Portal hypertension may also follow thrombosis of the splenic vein from pancreatitis or tumour. In this instance the portal hypertension is left sided (sectorial) and the varices affect the short gastric and gastroepiploic veins.

Obstruction to portal venous flow is followed by enlargement of natural portosystemic communications (Figure 24.6) and by the development of new collateral channels at surgically constructed mucocutaneous junctions (colostomy, ileostomy). Rarely, portal venous blood is shunted away from the liver by an enlargement of the umbilical vein (Cruveilhier–Baumgarten syndrome) and may be detected by a venous bruit in the midline. Though there is a risk of variceal bleeding from the ileum, colon and haemorrhoidal areas, the major risk of haemorrhage is from the oesophagus and stomach. In the oesophagus the varices are large, tortuous and thin walled with a tendency to rupture. However, in the stomach there is venous engorgement of the gastric mucosa with a tendency to erosive gastritis and a widespread diffuse haemorrhage; there is an additional component of proliferative growth of tiny blood vessels, which together cause the appearance of portal hypertensive gastropathy. The predilection of the gastric cardia to develop varices is probably due to the drainage of the left coronary vein after portal hypertension has developed. Instead of draining towards the liver, blood passes along paraoesophageal veins and then via 'perforator' veins to the submucosa of the oesophagus. Three columns tend to develop and run upwards for a variable length, usually communicating with the azygos system. Blood flow from the spleen may course through the short gastric vessels to the gastric fundus and link with enlarged collaterals at the cardia. Fundal varices are only detectable on retroflexion of the flexible endoscope in the stomach. Colonic varices may occasionally be seen on sigmoidoscopy, but not as commonly as may be expected. Nor is the caput medusa or periumbilical plexus of veins at all common. It is most prominent in patients who develop a small paraumbilical hernia into which an omental plug provides the portal flow.

It has to be recognized that not all patients develop portal hypertension as a result of their chronic liver disease; estimates vary from 15% to 40%. Furthermore, only about one-third of the patients with gastro-oesophageal varices (GOV) ever suffer from gastrointestinal bleeding. This is due both to the age of the patients and to the natural history of the underlying liver disease. In those who do develop portal hypertension, they may present with any combination of four clinical syndromes that can be attributed to portal hypertension:

- hypersplenism
- gastrointestinal bleeding
- ascites
- hepatic encephalopathy.

Hypersplenism

Splenomegaly is frequently associated with portal hypertension; it is due not just to passive venous hypertensive congestion but also to a hypertrophy of splenic tissue. This may lead to sufficient sequestration of formed blood elements in the spleen to cause haemolytic anaemia, leucopenia and thrombocytopenia. Only rarely are these features sufficient to produce major symptoms but they do lead to general debility. After portal vein decompression, hypersplenism improves in approximately 50% of patients.

Gastrointestinal haemorrhage

Approximately 70% of cirrhotics will have GOV at disease presentation. Of these patients, 30% will suffer gastrointestinal haemorrhage within 1 year of diagnosis; the index variceal bleed is associated with a 30–50% mortality rate. Bleeding risk is elevated in patients with larger varices, high portal venous pressure (>12 mmHg) and severe liver disease. The prognosis of a bleed is most dependent on the stage of liver disease prior to haemorrhage and to a lesser but important extent on the duration of the bleed and any hypotension, as this exacerbates decompensation of liver function. The following endoscopic findings are predictive of an increased risk of variceal haemorrhage:

- cherry-red spots
- red wale markings
- blue varices (as opposed to white).

Variceal bleeding may present with frank haematemesis, coffee-ground vomitus and/or repeated melaena. Management of oesophageal varices is divided into treatment of acute bleeding, prevention of recurrent haemorrhage and primary prophylaxis.

Management of acute variceal haemorrhage

Patients should be resuscitated with intravenous fluids and blood products as appropriate; those with known liver disease or stigmata of cirrhosis should receive vasoactive drug therapy (terlipressin or octreotide). Blood volume resuscitation should aim to restore haemodynamic stability but must be undertaken with caution; the recommended target haemoglobin is 8 g/dL. Experimental studies have shown that full restitution of lost blood volume results in an elevation of portal pressure above baseline, with increased risk of rebleeding and mortality. Upper gastrointestinal endoscopy should be performed as soon as possible, ideally after the patient has been resuscitated and is haemodynamically stable. On occasion, even with aggressive fluid resuscitation, stability may fluctuate and early endoscopy may stabilize the situation, but this should not slow the resuscitation. Endoscopy should be performed within 12–18 hours of admission.

Endoscopic therapy

Oesophageal varices

Endoscopy has the advantage of allowing specific therapy at the time of diagnosis; it is also important to exclude bleeding from other sources (e.g. peptic ulceration). Variceal haemorrhage may be treated endoscopically, either by injection sclerotherapy or by band ligation. Sclerotherapy involves injecting a sclerosant solution such as 5% ethanolamine oleate directly into the varix, producing vessel thrombosis, or into the overlying submucosa to induce inflammation and subsequent fibrosis. Complications include fever (transient), ulceration, stricture, perforation (rare), chest pain, mediastinitis and pleural effusion.

Endoscopic variceal ligation (EVL) is achieved by attaching a banding device to the tip of the endoscope. The varix is aspirated into the banding chamber and the band placed over the varix using a trip-wire mechanism, ligating the vessel. One to three bands are applied to each varix, resulting in thrombosis. Chest pain or banding-related ulceration may occur following EVL, but there are fewer associated complications than injection sclerotherapy. It has been demonstrated in a meta-analysis of 10 randomized controlled trials (RCTs) that EVL is superior to sclerotherapy in terms of both rebleeding rates and mortality; injection sclerotherapy had higher complication rates, particularly with regards to oesophageal ulceration and sepsis.

Combined endoscopic (either EVL or injection sclerotherapy) and pharmacological (splanchnic vasoconstrictors) therapy has proven superior to endotherapy alone in a meta-analysis of eight trials, with more effective control of initial haemorrhage and 5 day haemostasis without differences in morbidity or mortality. Both the British Society of Gastroenterology and the American Association for the Study of Liver Diseases (AASLD) recommend combination therapy with initiation of vasoactive drugs at the time of admission, prior to diagnostic upper gastrointestinal endoscopy.

Gastric varices

Bleeding from gastric varices only accounts for 10–36% of all variceal haemorrhages, but the bleeding can be more severe and the management more challenging. Gastric varices are classified according to their distribution and whether they are in continuity with oesophageal varices. GOV type 1 (GOV 1), the most common, are in continuity with oesophageal varices, extending <5 cm along the lesser curve. GOV which are in continuity with oesophageal varices but which extend further towards the fundus are classified as GOV type 2 (GOV 2). Isolated gastric varices (IGVs) occur in the absence of oesophageal varices: type 1 are located in the fundus (IGV 1) and type 2 (IVG 2) are located in the body, antrum or around the pylorus. GOV 1, GOV 2 and IGV 1 are most commonly associated with upper gastrointestinal haemorrhage.

Fewer clinical trials exist regarding endotherapy for gastric variceal haemorrhage, thus it is difficult to formulate management guidelines in the absence of robust evidence. GOV 1 constitute extensions of oesophageal varices and the recommended management is the same as for oesophageal varices. Haemostasis and rebleeding rates are reportedly similar in GOV 1 and oesophageal variceal haemorrhage.

Varices in the cardia and fundus (GOV 2 and IGV 1) tend to be more tortuous and complex; management of haemorrhage from these varices is more challenging and requires different endotherapy to GOV 1. Injection sclerotherapy is not recommended for gastric variceal bleeding following reports of inadequate haemostasis, embolization of sclerosant and ulcer haemorrhage at the site of the injection. Alternative therapeutic options are EVL or variceal obturation by injection of tissue adhesive glue or thrombin. Endoscopic variceal obturation (EVO) using cyanoacrylate injection is the current recommended treatment for fundal variceal haemorrhage, providing better control of initial haemorrhage and lower rebleeding rates than EVL. Two RCTs compared outcomes of EVO and EVL in the management of gastric variceal haemorrhage: the first found EVO to be superior both in the achievement of initial haemostasis and in lowering the rates of repeat haemorrhage; the second found no difference in control of bleeding but demonstrated a significant reduction in recurrent haemorrhage in patients treated with EVO. These studies, however, reported outcomes from all patients presenting with gastric variceal haemorrhage, including those with GOV 1. GOV 1 varices, as reported previously, are as well controlled by EVL as oesophageal varices, making these results difficult to interpret. Future studies limited to the management of GOV 2 and IGV 1 variceal haemorrhage are required. Thromboembolic complications including portal vein embolization (PVE), splenic infarction, myocardial infarction and stroke have been reported in case studies and a case series reported non-fatal pulmonary emboli in 4.6% of patients.

Endotherapy using thrombin injection has also been described in the control of gastric variceal haemorrhage but, as yet, this has not been subjected to RCT. Two small trials (patient numbers of 12 and 13) reported haemostasis rates of 100% and 92% and rebleeding rates of 27% and 0%, respectively. Both studies were limited by small patient numbers and short-term follow-up. Potential risks of thrombin endotherapy include allergic reactions or thromboembolic complications.

Vasoconstrictor therapy

Vasoactive drugs restrict portal inflow by splanchnic arterial vasoconstriction, resulting in reduced portal pressure. Therapy should be initiated as soon as variceal haemorrhage is suspected, prior to endoscopy. The most commonly used vasoactive drugs are discussed below.

- *Octreotide* – a somatostatin analogue. Administered as a 50 μg bolus followed by 50 μg/h continuous infusion. Produces splanchnic vasoconstriction without significant systemic vascular effect or complications.
- *Terlipressin* – a synthetic analogue of vasopressin. Administered at an initial bolus dose of 2 mg every 4 hours, titrated down to 1 mg every 6 hours after haemostasis is achieved. Contraindicated in patients with ischaemic heart disease, terlipressin has longer biological activity and fewer side effects than vasopressin (e.g. arrhythmias, hypertension, digital, gastrointestinal or cardiac ischaemia).

A meta-analysis of seven RCTs showed a statistically significant reduction in mortality in patients treated with terlipressin compared with placebo; the number needed to treat to prevent one death was 8.3. Combined vasoconstrictor and endoscopic therapy is superior to endotherapy alone; terlipressin combined with endoscopy showed improved haemostasis and reduced mortality compared with endotherapy plus placebo. UK treatment guidelines recommend initiation of terlipressin therapy in all patients with suspected variceal haemorrhage; treatment should be continued for 48 hours after confirmation of diagnosis.

Terlipressin is not yet available for treatment of variceal haemorrhage in North America; somatostatin and its analogues, most commonly octreotide, are administered as recommended by the AASLD. A systemic review identified 21 RCTs studying somatostatin and its analogues; there was no change in mortality but improved rates of initial haemostasis were demonstrated with drug therapy compared with placebo. A meta-analysis of eight RCTs combining endoscopy with somatostatin and its analogues showed combination treatment was superior to endotherapy alone; initial haemostasis and rebleeding rates were significantly lower with combination therapy but there was no survival benefit. Patients should receive somatostatin and its analogues at initial presentation with variceal bleeding; treatment should continue for 3–5 days following endotherapy.

In summary, combined therapy should be the standard of care for acute variceal bleeding with an immediate pharmacological agent when variceal bleeding is suspected; where available terlipressin is probably the agent of choice and this should be combined with early endoscopy and therapy with EVL for oesophageal varices and EVO for gastric varices.

Management of uncontrolled variceal haemorrhage

Despite urgent pharmacological and endoscopic therapy, uncontrolled haemorrhage or early rebleeding occurs in 10–20% of patients. Initial emergency measures to control massive blood loss should be followed by definitive second-line management of the underlying cause.

Balloon tamponade

This procedure is a temporary salvage treatment for uncontrolled variceal bleeding, two balloon systems are commonly used.

- *Sengstaken–Blakemore tube* – with a gastric balloon of about 60–100 mL capacity, which is meant to anchor the tube at the cardia and an oesophageal balloon which is inflated to 30–40 mmHg and compresses the submucosal veins of the oesophagus.
- *Linton tube* – with a large gastric balloon of 300–700 mL capacity which is pulled against the diaphragmatic hiatus with traction and disconnects the high-pressure portal venous system from the thoracic azygous veins.

Haemostasis with balloon tamponade is achieved in 80–95% of patients with oesophageal or gastric variceal haemorrhage; it is, however, associated with serious complications such as aspiration pneumonia, mucosal ulceration and oesophageal rupture, which lead to mortality rates of 5–20%. As the use of the oesophageal balloon increases the risk of complications with little evidence of additional haemostatic effect, we should discourage its routine inflation. Haemorrhage of the severity to warrant use of a tamponade tube should also prompt consideration of elective tracheal intubation to prevent aspiration. Balloon tamponade should only be used as a temporizing measure (12–24 hours) in patients with uncontrolled haemorrhage; definitive endotherapy or portosystemic shunting can be performed once the patient is stabilized.

Transjugular intrahepatic portocaval shunt

Transjugular intrahepatic portocaval shunt (TIPS) is performed by interventional radiology and creates a channel between the systemic and portal venous systems to divert portal flow into the hepatic vein, thus reducing portal pressure. A guide wire is passed via the internal jugular vein into the right hepatic vein; the wire then passes via a needle through the liver parenchyma into the portal vein. A metal stent is then deployed across the intrahepatic parenchymal tract to form a shunt between the portal and hepatic veins. Procedure-related complications include pneumothorax, puncture site haematoma or arteriovenous fistula formation, cardiac arrhythmias, haemorrhage, arterial and bile duct injury. Postprocedure complications include hepatic encephalopathy, shunt-related haemolysis and stenosis or occlusion of the shunt. Acute alterations in cardiac output and central venous and pulmonary wedge pressure immediately post-TIPS insertion may result in cardiac failure or pulmonary oedema and patients must be monitored closely for these complications.

TIPS is safer for patients with advanced liver disease, avoiding the risks of general anaesthesia and major surgery. A large retrospective study comparing emergency TIPS insertion to oesophageal transection and devascularization showed improved survival with TIPS (42% vs 79%) and recommended this should be the treatment of choice for torrential variceal haemorrhage. Clinicians tend to have a lower threshold for TIPS insertion in gastric variceal haemorrhage: TIPS is recommended after a single episode of rebleeding or in cases where endotherapy is not possible. TIPS achieves haemostasis rates in excess of 90% in patients with uncontrolled gastric variceal haemorrhage.

Additional management of acute variceal haemorrhage

- *Antibiotic prophylaxis* – chronic liver disease patients with acute upper gastrointestinal haemorrhage have a high risk of developing serious bacterial infections such as spontaneous bacterial peritonitis (SBP); severe sepsis is associated with early recurrence of variceal bleeding and increased mortality. A meta-analysis of 12 trials evaluating antibiotic prophylaxis in cirrhotic patients with upper gastrointestinal bleeding demonstrated a significant reduction in infection rates, rebleeding and overall mortality. Short-term (7 days) oral norfloxacin or intravenous ceftriaxone is recommended.
- *Sucralfate* – 1 g four times daily is given to prevent mucosal ulceration following EVL or sclerotherapy.
- *Measures to prevent encephalopathy* – PSE may be precipitated by gastrointestinal haemorrhage; the management is described later in this chapter.

Prevention of recurrent variceal haemorrhage

Following the index episode of variceal bleeding, the risk of recurrence is 60–80% within 1–2 years with a mortality of 20–33%. Secondary preventative measures should be commenced prior to discharge from hospital.

Endoscopic therapy

Repeated courses of endotherapy (at weekly intervals) leads to obliteration of oesophageal varices by fibrous tissue, this usually requires two to four sessions. Once eradicated, patients should undergo surveillance endoscopy every 3–6 months to screen for variceal recurrence. Band ligation (EVL) has been shown to be superior to sclerotherapy in the prevention of variceal recurrence; a meta-analysis of seven trials showed reduced rebleeding rates, mortality and complications with EVL. Complications occur in approximately 14% of patients but are relatively minor, with transient dysphagia and chest discomfort occurring most frequently. Ulceration at the site of EVL is common and treatment with sucralfate and/or esomeprazole is recommended to reduce the risk of banding ulceration and bleeding.

There is no current evidence to support repeat courses of endotherapy in patients with gastric varices. A single RCT compared histoacryl glue obliteration with propranolol secondary prevention and found no difference in rebleeding or mortality rates; endotherapy was associated with more complications than beta-blockade.

Vasoactive drug therapy

Propranolol monotherapy was compared to placebo secondary prevention in a meta-analysis of 12 trials and shown to significantly lower the risk of rebleeding and reduce mortality. Combined beta-blockade and endotherapy was demonstrated to be superior to EVL alone in two RCTs, with rebleeding rates of 23% and 14%, respectively, for nadolol plus EVL compared with 47% and 38% for EVL alone. A combination of beta-blocker and nitrate has been shown to be superior to beta-blockade monotherapy and of equal efficacy to EVL. In practice, however, few patients are able to tolerate dual pharmacotherapy because of side effects, and the current recommended practice is combined beta-blockade plus EVL for secondary prevention.

Transjugular intrahepatic portocaval shunt

The use of TIPS for secondary prevention is currently reserved for patients in whom endotherapy and/or pharmacological therapy has failed or cannot be tolerated. A meta-analysis of 22 trials comparing portosystemic shunts (TIPS and surgical shunts) with endotherapy demonstrated a reduction in recurrent haemorrhage, but this was at the expense of an increased incidence of chronic encephalopathy with no survival benefit.

Standard practice has recently been challenged by researchers in Barcelona who argued that patients with advanced liver disease and high risk of rebleeding would benefit from early TIPS insertion. Patients with Child's C cirrhosis or Child's B disease with active bleeding at endoscopy were randomized to receive standard medical secondary prevention or early TIPS insertion within 72 hours of admission. The early TIPS group showed significantly lower rates of treatment failure, encephalopathy and mortality than patients managed with EVL and vasoactive drug therapy. Use of more modern polytetrafluoroethylene -covered stents is associated with lower risk of encephalopathy and improved rates in patency and rebleeding. The study is limited, however, by selection bias (e.g. exclusion of patients with Child–Pugh score >13) and low patient numbers.

Primary prophylaxis

All patients with cirrhosis should undergo endoscopic screening to detect varices at the time of diagnosis, and if not present at 3 year intervals thereafter. Beta-blockade is recommended in patients with larger varices or in those with small varices and additional risk factors for haemorrhage (e.g. Child's B/C liver disease, presence of red wale marks on varices). EVL has been shown to be as effective as beta-blockade in primary prophylaxis, but much less cost-effective so is reserved for those intolerant of beta-blockade.

Surgical treatment

More than 85% of patients with bleeding varices can be controlled by non-surgical measures and surgical treatment of refractory variceal bleedings has become even more unusual since the introduction of TIPS and of prophylactic endoscopic sclerotherapy.

However patients in whom endoscopic sclerotherapy has failed or the bleeding has recurred early remain problematic. The majority of these patients are high-risk candidates with poor liver function from end-stage liver disease or where liver function has worsened from Child A/B to C as a consequence of the recurrent hypovolaemic episodes. Unless prior information is available on the patient, it is difficult to distinguish between these two categories at the time of emergency presentation. Portosystemic shunting in this group carries a prohibitive mortality and is contraindicated. The aim of surgical management should be to prevent death from exsanguination and not to treat the underlying portal hypertension and therefore an oesophageal transection should be considered instead. This procedure was proposed by Sugiura and Futagawa and consisted of oesophageal transection, extensive oesophagogastric devascularization and splenectomy. Nowadays, the transection and reanastomosis of the oesophagus is performed with a circular stapler introduced through a service gastrotomy (Figure 24.17).

Some other high-risk patients can be considered for surgical treatment of portal hypertension in an elective setting, such as those in whom variceal bleeding has been temporarily arrested and who have good liver function (Child A/'good' B) but remain at high risk of rebleeding and have had several episodes of sclerotherapy. Elective surgical therapy by portosystemic shunting or oesophageal transection with devascularization may be a sensible, cost-effective and safe management approach.

Patients with end-stage liver disease (Child C) who have been definitively controlled by sclerotherapy should be considered

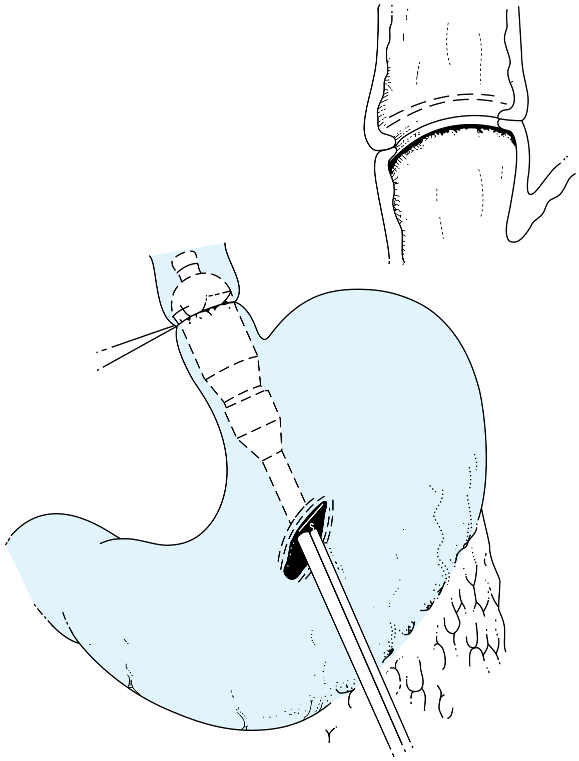


Figure 24.17 Stapled transection and reanastomosis of the distal oesophagus with a circular stapler (28–31 mm) introduced through a small anterior gastrotomy during a Sugiura procedure. A strong suture is tied around the central axle of the stapler gun before it is closed and fired.

for hepatic transplantation. For this reason, surgery should be avoided as it increases the morbidity and mortality of the procedure and TIPSs are more appropriate and successful in achieving portal decompression and thus avoid further bleeding episodes prior to hepatic transplantation.

Portosystemic shunts can be total (total deviation of portal blood flow to the systemic circulation), partial (partial deviation of portal blood flow though the placement of a narrow prosthesis in either the portocaval or mesocaval position) and selective (distal splenorenal, coronariocaval, splenocaval). Selective shunt operations are associated with good control of rebleeding and lower rates of encephalopathy but are less successful in controlling acute variceal bleeds. The indications for such an operation are now limited to non-cirrhotic portal hypertension and patients living in areas where newer therapies (TIPS) are not available. The most commonly performed of the selective shunts, and one that still has a role to play nowadays, albeit infrequently, is the distal splenorenal shunt (DSRS or Warren shunt). In this operation (Figure 24.18), the splenic vein is anastomosed to the left renal vein in an end-to-end fashion with diversion of blood from the cardia via the short gastric veins into the inferior vena cava circulation. The objective is to preserve portal venous flow to the liver and while this is achieved in the postoperative period, it does not seem that it can be maintained in the long term. A multicentre randomized trial of TIPS vs DSRS showed no overall difference in survival and a tendency for TIPS to be more cost-effective in terms of lives saved.

Ascites

Ascites is the commonest complication of hepatic cirrhosis, occurring in 50–60% of patients within 10 years. Ascites is a major cause of hospital admission and its development is an important landmark in the natural history of cirrhosis, being associated with 50% mortality over 2 years. Accumulation of ascites occurs with the development of severe portal hypertension, impaired sodium excretion and water retention. The pathogenesis of ascites has several theories and is not yet fully understood; these have been discussed in great detail in a number of review articles. This chapter will briefly discuss the most accepted mechanisms, the peripheral arterial vasodilatation hypothesis and the forward theory of ascites formation, on which current treatment recommendations are based.

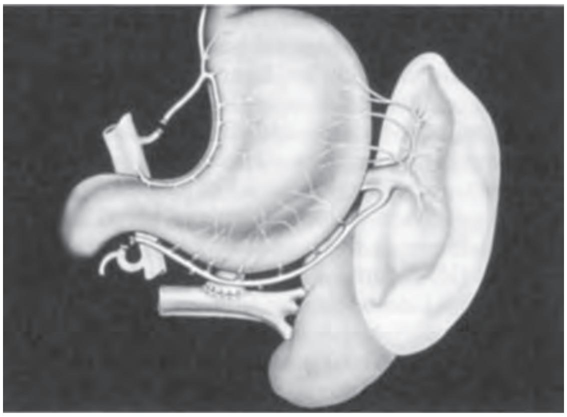


Figure 24.18 The Warren distal splenorenal shunt which results in transplenic decompression of the portal circulation.

The presence of portal hypertension is essential for the development of ascites; it usually occurs when the wedged hepatic venous portal gradient exceeds 12 mmHg but can occur at lower pressures. Portal hypertension disrupts the Starling equilibrium within the splanchnic circulation, resulting in transudation of fluid into the peritoneal cavity.

Peripheral arterial vasodilatation hypothesis

This theory postulates that renal sodium retention occurs as a result of splanchnic arterial vasodilatation secondary to portal hypertension. Nitric oxide is considered to be the putative vasodilator, although other substances (e.g. atrial natriuretic peptide, glucagon calcitonin gene-related peptide and prostaglandins) may also be involved. The reduction in effective arterial blood volume activates the sympathetic nervous system and renin–angiotensin system, thus promoting sodium and water retention. The resultant increase in extracellular fluid volume leads to formation of ascites and oedema.

Forward theory of ascites formation

According to this theory, splanchnic arterial vasodilatation induces ascites formation by simultaneously impairing the systemic circulation, leading to sodium and water retention, and splanchnic microcirculation, resulting in leakage of fluid into the peritoneal cavity.

Assessment and diagnosis

The initial evaluation of the patient presenting with ascites includes a detailed history, physical examination, abdominal ultrasound, laboratory assessment of hepatic and renal function, urine and serum electrolytes and ascitic fluid analysis. Diagnostic paracentesis of 10–20 mL of fluid should be obtained and the following analyses performed:

- cell count – a neutrophil count >250 cells/mm^3 is diagnostic of SBP
- Gram stain and culture
- protein and albumin – the serum–ascites albumin gradient (SAAG) helps differentiate ascites due to portal hypertension from ascites due to other causes; a SAAG 11 g/L is ascribed to portal hypertension with 97% accuracy; total ascitic protein should be measured to assess the risk of SBP, patients with protein <10 g/L have a higher infection risk.
- cytology – to assess for malignant cells
- amylase – to exclude pancreatic ascites.

Management

Treatment of ascites aims to restrict dietary salt intake and increase renal sodium excretion; patients with tense ascites will require paracentesis prior to commencement of medical therapy.

Sodium restriction

Negative sodium balance can be achieved by dietary salt restriction alone in 10–20% of patients; in practice it is usually combined with diuretic therapy during initial management. Sodium restriction [2000 mg/day (88 mmol/day)] is associated with reduced diuretic requirements, more rapid resolution of ascites and shorter hospital stay. Recommended sodium intake for ascitic patients varies between guidelines from 2 to 6.9 g/day; in practice, patients are advised to follow a no added salt diet and to avoid processed foods. This achieves reasonable sodium restriction without impairing nutrition.

Diuretics

Spironolactone, an aldosterone antagonist, acts on the distal tubules to promote natriuresis and conserve potassium. It is the drug of choice for the initial management of cirrhotic ascites and is commenced at 100 mg daily, increasing to a maximum dose of 400 mg. It takes 3–5 days from onset of therapy for the drug to take effect; spironolactone should be increased in 100 mg increments every 3 days until adequate natriuresis has been achieved. The aim of diuretic therapy is to increase urinary sodium excretion so that it exceeds 78 mmol/day (88 mmol intake/day to 10 mmol non-urinary excretion/day). A 'spot' urine specimen with sodium concentration more than potassium concentration correlates with a 24-hour sodium excretion greater than 78 mmol/day with approximately 90% accuracy. Random testing of the urine sodium–potassium ratio is more convenient than the cumbersome 24-hour collection, and gives an indication of whether diuretic dosage is adequate or requires further titration. Common side effects of spironolactone are hyperkalaemia, renal impairment and those related to the drug's antiandrogenic activity (e.g. reduced libido, gynaecomastia). Amiloride, 10–40 mg daily, may be substituted for spironolactone in patients with tender gynaecomastia, but it has been found to be less effective and is more expensive than spironolactone.

Furosemide, a loop diuretic, causes marked diuresis and natriuresis in the normal population but has been shown to be less effective in cirrhosis. It is used as an adjunct to spironolactone therapy with initial dosing of 40 mg daily, titrating to a maximum 160 mg. Higher doses are associated with electrolyte disturbance and metabolic alkalosis. All patients should be monitored closely in the initial stages of management; overdiuresis may result in volume depletion, renal impairment hyponatraemia or PSE.

Spironolactone monotherapy is useful in patients with minimal fluid overload who can be managed in the outpatient setting. Combination therapy is recommended for patients with more marked fluid retention, resulting in faster fluid mobilization and maintenance of normokalaemia.

Refractory ascites

Refractory ascites occurs in patients who fail to respond to sodium restriction and diuretics (diuretic-resistant ascites) or in those who develop complications which preclude the use of effective doses of diuretics (diuretic-intractable ascites). The prognosis is poor in such patients and, if appropriate, referral for liver transplantation should be considered. Other treatment options include paracentesis and portosystemic shunts.

Paracentesis

Patients with refractory ascites are initially managed by serial large volume paracentesis (LVP); total paracentesis is generally safer than repeated smaller aspirations. Volume expansion is given alongside paracentesis to avoid postprocedure circulatory dysfunction, renal impairment and electrolyte disturbance. Several studies have demonstrated the safety and effectiveness of LVP with volume expansion; administration of 8 g albumin/L ascites removed (100 mL 20% albumin/3 L ascites) is recommended.

Transjugular intrahepatic portocaval shunt

TIPS insertion should be considered in patients in whom frequent LVPs are required; prospective RCTs have shown TIPS to be more effective in controlling ascites than LVP, but there is no difference in survival. Hepatic encephalopathy occurs following TIPS insertion in 25% of patients, with a higher risk in patients over 60. It may precipitate pulmonary oedema and cardiac failure in patients with pre-existing heart disease because of increased cardiac preload. Outcomes are poorer in patients with advanced (Child's C liver disease), and TIPS is not recommended in this group. It is also advisable to avoid TIPS in patients with recurrent hepatic encephalopathy, concomitant infection, progressive renal failure or severe cardiopulmonary disease.

Spontaneous bacterial peritonitis

SBP is a monomicrobial infection of ascites occurring in 10–30% of hospitalized patients with cirrhosis. It is a serious complication of ascites with a mortality rate of 20%, despite improvements in early diagnosis and prompt treatment. Patients with SBP are frequently asymptomatic; therefore, diagnostic paracentesis is mandatory in all patients presenting to hospital with ascites and in those with worsening decompensation.

Escherichia coli, Gram-positive cocci (mainly streptococci) and enterococci are the most commonly isolated organisms, accounting for 70% of cases of SBP. Intravenous cefotaxime covers 95% of the flora commonly isolated in SBP and achieves high penetrance of ascitic fluid during therapy. Empirical treatment with cefotaxime is recommended in all patients with ascitic neutrophil count >250 cells/mm^3 until results from culture and antibiotic sensitivity testing are available. Patients with a negative cell count should receive antibiotic prophylaxis, norfloxacin 400 mg daily, while in hospital.

Hepatorenal syndrome

Hepatorenal syndrome (HRS) is defined as the occurrence of renal failure in patients with advanced liver disease in the absence of an identifiable cause of renal impairment (e.g. hypovolaemia, nephrotoxic drugs). HRS is a functional renal impairment resulting from renal hypoperfusion, the pathophysiology is similar to that producing ascites. There are two types:

- type 1 HRS – characterized by a rapid and progressive impairment in renal function (increase in serum creatinine 100% from baseline to >221 mol/L within 2 weeks)
- type 2 HRS – occurs in patients with refractory ascites, characterized by a stable or less progressive impairment in renal function.

Type 1 HRS commonly affects patients with severe alcoholic hepatitis or those with end-stage liver disease following an episode of sepsis such as SBP. Development of type 1 HRS has an extremely poor prognosis with 80% mortality within 2 weeks; management involves plasma volume expansion with albumin and vasoconstrictor therapy. Liver transplantation should be considered in all patients with HRS, as post liver transplantation survival is approximately 65% for patients with type 1; optimization of renal function preoperatively improves outcomes after liver transplantation. Terlipressin and albumin therapy has been shown to improve renal function in 50–60% of patients; however, studies to date show no improvement in long-term survival without liver transplantation. TIPS may offer an alternative to pharmacological therapy but there are insufficient data to support this at present; haemodialysis is currently reserved for patients awaiting transplantation who have failed to respond to medical therapy.

Renal dysfunction in patients with cirrhosis

The classification of renal dysfunction in cirrhotics has recently been expanded, recognizing that many patients develop acute and chronic renal dysfunction without the diagnostic criteria of HRS. Early diagnosis of lesser degrees of renal impairment should improve outcomes as patients will be treated earlier in the natural history of renal dysfunction. The revised classification system has been expanded to include:

- acute kidney injury – rise in serum creatinine of ³50% from baseline or by 26.4 mol/L in less than 48 hours; HRS type 1 is a specific form of acute kidney injury
- chronic kidney disease – glomerular filtration rate (GFR) <60 mL/min for >3 months; HRS type 2 is a specific form of chronic kidney disease
- acute on chronic kidney disease – rise in serum creatinine of ³50% from baseline or by 26.4 mol/L in less than 48 hours in a cirrhotic patient whose GFR is <60 mL/min for more than 3 months.

Portosystemic encephalopathy

PSE is defined as a spectrum of neuropsychiatric disorders in patients suffering from liver disease after exclusion of other known neurological disease. It is characterized by personality changes, intellectual impairment, reduced level of consciousness and abnormalities on psychometric testing. PSE occurs when blood bypasses the liver via portosystemic collaterals, or following portosystemic shunt procedures such as TIPS, and neurotoxic substances cross the blood–brain barrier. It most commonly occurs in the setting of cirrhosis and portal hypertension but a similar acute encephalopathy occurs in FHF.

The encephalopathy in chronic liver disease may be subclinical or overt, either as recurrent reversible episodes or as an overt and persistent condition when PSE may be accompanied by permanent structural central nervous system changes, predominantly cortical atrophy of the brain. Acute onset encephalopathy often has a precipitating factor or may reflect ongoing progressive liver failure: the patient becomes increasingly drowsy, eventually becoming comatose. Factors that may precipitate PSE include infection, gastrointestinal haemorrhage, constipation, central nervous system depressant drugs, TIPS insertion, surgical procedures and fluid and electrolyte disturbance due to paracentesis or diuretics. The acute encephalopathy that occurs in FHF shares many of the pathogenic mechanisms as that of cirrhosis, but cerebral oedema plays a more prominent role.

Pathogenesis

Despite considerable research, the exact mechanisms involved (probably multifactorial) in the pathogenesis of PSE remain uncertain. Animal and *in vitro* models have found evidence of astrocyte swelling and oxidative stress and have demonstrated abnormalities in glutaminergic, serotoninergic, -aminobutyric acidergic (GABA) and catecholamine pathways. Most theories accept that nitrogenous substances derived from the gut adversely affect brain function, producing alterations in neurotransmission that alter consciousness and behaviour. Hyperammonaemia is found in most patients with PSE and reductions in blood ammonia levels are associated with clinical improvement; there is, however, no correlation between laboratory ammonia levels and severity of encephalopathy. In addition there is now mounting evidence for microglial inflammation having a pathogenic role in encephalopathy.

Management

The principles of management of patients with PSE include the search for and correction of any precipitating cause, e.g. withdrawal of sedatives, correction of fluid and electrolyte disturbances, revision or closure of portosystemic shunts, arrest of haemorrhage and treatment of sepsis. Other measures aim to provide supportive care and reduce production and absorption of nitrogenous substances from the gut.

Contrary to widespread belief, patients with chronic liver disease should not be subjected to dietary protein restriction. In the first instance, patients with stable cirrhosis have a higher protein requirement than normal, i.e. 1.2–1.5 g/kg body weight to remain in positive balance. The European Society for Parenteral and Enteral Nutrition recommends a daily protein intake of 1.0–1.5 g/kg depending on the degree of hepatic decompensation. In the rare patients unable to tolerate this amount due to PSE, protein intake is reduced to 0.5 g/kg and the remainder of their requirements is achieved by administration of branched chain amino acids. In practice, the most effective way of achieving a sufficient protein intake in patients with liver disease is by frequent small meals and a late evening meal. This has been shown to improve nitrogen balance without exacerbating encephalopathy.

Non-absorbable disaccharides such as lactulose inhibit intestinal ammonia production by a number of mechanisms. Conversion of lactulose to lactate in the caecum reduces intraluminal pH, inhibiting the growth of ammoniagenic intestinal bacteria; the acidic pH also alters the metabolism of the intestinal flora, promoting uptake of ammonia. Lactulose also has a cathartic effect; the dose should be titrated to produce two to four soft bowel movements daily.

Antibiotics such as neomycin, metronidazole, vancomycin and rifaximin may be administered to reduce ammonia production by intestinal bacteria. Their use is limited by adverse effects including ototoxicity, renal impairment and peripheral neuropathy. Rifaximin, however, has been shown to be as effective at treating PSE as other antibiotics and is better tolerated than both cathartics and other non-absorbable antibiotics.

Liver transplantation

Advances in the field of liver transplantation over the past few decades have resulted in dramatic improvement in outcomes, with 1-year survival rates of 80–95% and 5-year survival rates of 62–80%. Transplantation should be considered in all patients presenting with acute FHF and in those suffering from chronic liver disease in whom life-threatening complications have developed.

Indications for liver transplantation

Acute liver failure

Patients with FHF from any cause should be referred to a transplant centre; hepatotoxicity due to acetaminophen toxicity or idiosyncratic drug reactions is the most common aetiology in the UK and USA.

Chronic end-stage liver disease

Irrespective of the underlying aetiology, development of the following complications warrants consideration of referral for liver transplantation:

- recurrent variceal haemorrhage
- refractory ascites
- SBP
- recurrent encephalopathy
- worsening synthetic dysfunction
- HCC

Special consideration must be given in the following cases:

- Alcohol-related cirrhosis – a 6-month period of abstinence prior to consideration for liver transplantation is recommended; in some cases liver function may improve to such an extent that liver transplantation is no longer required.
- Chronic hepatitis B – recurrence of HBV infection in the graft can be prevented by antiviral therapy pre-liver transplantation and continued treatment post-liver transplantation along with HBV immunoglobulin prophylaxis.
- Chronic hepatitis C – synthetic dysfunction, cytopenias and/or reduced tolerance to side effects often precludes antiviral therapy in end-stage HCV cirrhosis and even if undertaken has a low chance of success. If still HCV RNA positive at the time of transplantation, infection of the graft post-liver transplantation is inevitable, with approximately 50% developing graft cirrhosis within 5 years.
- HCC – disease recurrence post-liver transplantation occurs in patients with larger tumours, mutifocal disease and lymphatic or vascular invasion; this is associated with a high mortality. Liver transplantation is only considered for patients with a single HCC <5 cm or ≤3 lesions with the largest <3 cm, although these criteria are being challenged, with arguments that the biological behaviour of the tumour also has prognostic impact on the outcome of transplantation.

Contraindications

Absolute contraindications to liver transplantation include active sepsis, metastatic or extrahepatic malignancy, cholangiocarcinoma, advanced cardiopulmonary disease and AIDS.

Relative contraindications include HIV infection, HBV DNA positivity, active substance misuse, severe psychiatric disorder, portal venous thrombosis, pulmonary hypertension and age >70 years. All cases should be considered on an individual basis and discussed with a transplant centre if deemed appropriate. These relative contraindications worsen outcomes but if other factors are favourable, then transplantation may be justified.

Patient selection and organ allocation

Unfortunately the number of patients requiring liver transplantation continues to exceed deceased donor organ availability; despite surgical innovations including use of extended criteria donors, split livers, donation after cardiac death and live liver donation, many patients still die on the transplant waiting list. Scoring systems to assess need for liver transplantation and prioritize organ allocation to patients with the highest risk of death without liver transplantation are employed to rationalize resources and reduce waiting list mortality.

Model for end-stage liver disease

The model for end-stage liver disease (MELD) score is based on three variables, (1) serum bilirubin, (2) serum creatinine and (3) international normalized ratio (INR), and has been shown in both retrospective and prospective studies to be a reliable predictor of 3 month mortality in end-stage cirrhosis. It is not used as criteria for listing but is used to prioritize listed patients for organ allocation, in the USA and across Europe with the exception of the UK.

United Kingdom model for end-stage liver disease

The United Kingdom MELD (UKELD) score is derived from the patient's serum sodium, creatinine, bilirubin and INR. Hyponatraemia in cirrhotic patients is associated with neurological dysfunction, refractory ascites, increased risk of development of HRS and a higher mortality. Given the important prognostic value of sodium, it has been incorporated as an adjunct to the MELD score in UK organ allocation.

Minimal listing criteria require that the patient should have projected 1 year liver disease mortality without transplantation of >9%. This is predicted by a UKELD score of 42. Patients should be followed closely after listing for liver transplantation and constantly re-evaluated with regard to their eligibility and fitness to undergo transplantation. Evaluation of patients assesses both physical and psychological health to ensure appropriate patient selection and optimal utilization of this scarce resource.

Hepatic abscesses

Abscesses of the liver are less common in temperate than in tropical regions. There are also differences in the underlying aetiology between the two regions. There are two types: pyogenic and amoebic.

Pyogenic abscesses

Pathogenesis

The incidence of pyogenic liver abscess has remained relatively constant over the past century despite earlier diagnosis and treatment of underlying causes and more aggressive antibiotic therapies. In recent years, the decrease in cases resulting from haematogenous spread from infected foci has been mirrored by an increase in cases secondary to hepatobiliary pathology. The main aetiological factor is bile duct infection with ascending cholangitis commonly due to *E. coli* and anaerobic Gram-negative organisms. In the UK, biliary sepsis will be the predisposing factor in almost half of the patients reviewed. Hepatic abscesses secondary to ascending cholangitis are often multiple due to the distribution of the infecting organism along the biliary ductal system. Early reports implicated choledocholithiasis as the main causative factor; however, more recent series document malignant biliary obstruction as a more common aetiological factor. Other sources of infection include the following:

- Ascending pylephlebitis. While any inflammatory process within the abdomen may initiate pylephlebitis, it is most commonly the result of complicated diverticulitis.
- Some hepatic abscesses of staphylococcal and streptococcal origin arise as a complication of bacteraemia (haematogenous).
- Direct extension from intra-abdominal suppuration, e.g. gangrenous cholecystitis, penetrating peptic ulcer disease and subphrenic collections.
- Trauma to the liver, both penetrating and non-penetrating, may devitalize liver tissue and subsequent infection produces an abscess.
- A significant group of patients are found in the elderly population. No obvious cause is found in many of these patients (cryptogenic). These often have an insidious onset and non-specific symptoms such that, at the time of diagnosis, the abscess is usually very large. In some cases, the chronic abscess erodes through the diaphragm and bursts into the bronchial tree presenting with chest symptoms (Figure 24.19). The infecting organisms for these abscesses are commonly the *Peptostreptococcus* and *Streptococcus milleri*, but other microbes including *Bacteroides fragilis* may be involved.
- Parasitic infestations – *Ascaris lumbricoides.*
- Tuberculous liver abscesses.
- HIV infection – cholangitis/cholangiopathy.

HIV infection is now one of the risk factors, and abscesses due to meticillin-resistant *Staphylococcus aureus* appear to be on the increase. The infection is polymicrobial in 45%. *E. coli* and *S. milleri* are the most frequently isolated organisms. In endemic areas of *Ascaris lumbricoides* infestations, e.g. Kashmir, India, up to 15% of pyogenic hepatic abscesses are associated with this infestation.

Cholangitis/cholangiopathy associated with HIV infection is characterized by chronic abdominal pain, low-grade fever, cholestasis, and sometimes areas of focal or diffuse dilatation of the bile ducts. The disease appears to be the result of immunosuppression and/or secondary opportunistic infections rather than a direct cytopathic effect of the virus itself. Various opportunistic pathogens, including cytomegalovirus, *Cryptosporidium, Campylobacter fetus* and *Candida albicans*, have been implicated in the aetiology of HIV-associated cholangitis.

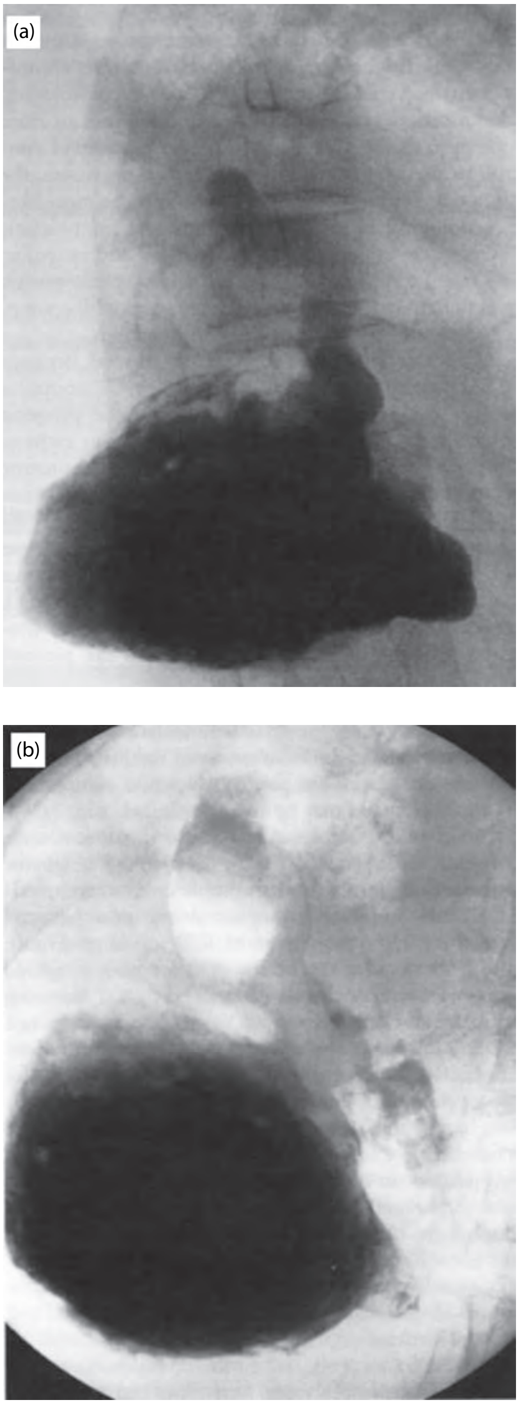


Figure 24.19 Contrast study showing a fistula between the right hepatic abscess and right basal bronchial tree in an elderly patient initially treated with percutaneous drainage.

Pylephlebitis usually occurs secondary to infection in the region drained by the portal venous system, the most common being diverticulitis and appendicitis. Infection is caused by *E. coli* in 54%, followed by *Proteus mirabilis* (23%). The pylephlebitis induces a septic thrombosis of the portal vein and its branches, with multiple microabscess formation in the liver (honeycomb liver). Early recognition of the disease and timely antibiotic therapy are essential for survival. The overall reported mortality is 32%.

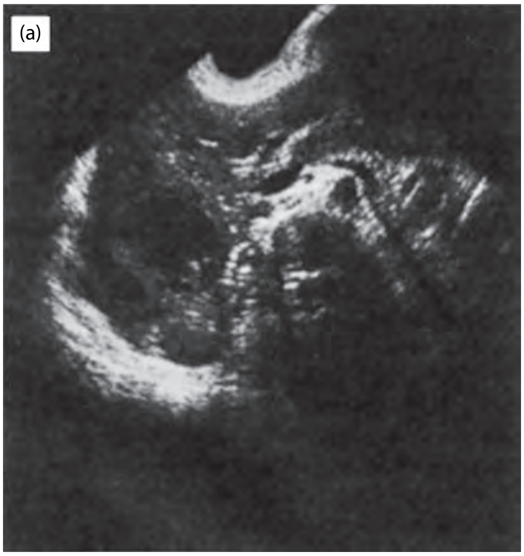
Irrespective of aetiology, liver abscesses are found much more commonly in the right lobe. All abscesses contain areas of liver parenchymal cell necrosis surrounded by polymorphonuclear leucocytes and lymphocyte infiltration with relatively damaged parenchyma and viable bacteria on the periphery. Ultimately a fibrous reaction is initiated which may produce a fibrous capsule containing pus.

Clinical features of hepatic abscesses

Untreated pyogenic liver abscesses carry a high mortality and 10% are diagnosed only at postmortem. The important determinants of mortality are multiple abscesses, hyperbilirubinaemia and comorbid disease. Since most pyogenic abscesses are secondary to other infective processes, the clinical features may be dominated by the primary disorder. Characteristically there is a high fever, rigors, profuse sweating, anorexia and vomiting, with pain as a relatively late symptom. In elderly people pyogenic liver abscess is commoner in women, and the symptoms may be vague. Thus the patients may have minimal abdominal tenderness on physical examination. The clinical features are also less striking in hepatic amoebiasis where the fever is usually low grade. However, pain is a more common feature with amoebic abscess and is aggravated by movement and coughing. An amoebic abscess may reach a very large size before causing pain if situated posteriorly and pointing to the bare area of the liver. About half the patients with amoebic abscesses will have diarrhoea.

Hepatomegaly is common, particularly with amoebiasis. Occasionally with right lobe abscesses there is bulging and pitting oedema of the intercostal spaces. An abscess in the left lobe may present as a painful epigastric swelling. On investigation, anaemia and leucocytosis may be found, and raised levels of acute phase proteins. Disturbances of LFTs are not diagnostic, particularly when complicated by cholangitis, and may be absent in amoebiasis.

Blood cultures are usually positive in patients with pyogenic abscesses when taken during the height of pyrexia. Both aerobic and anaerobic cultures are needed. Clinical suspicion of hepatic abscess must be confirmed by ultrasonography or CT scanning (Figure 24.20). The highest diagnostic yield for both pyogenic and amoebic liver abscess (ALA) is obtained by CT scanning with contrast enhancement. Magnetic resonance cholangiography should be undertaken in patients with biliary symptoms, obstructive LFTs or a dilated common bile duct, and can be combined with cross-sectional MRI to identify any hepatitic parenchymal abnormality. Diagnostic aspiration is a safe and reliable procedure and provides means of identification of the organisms responsible and thus the appropriate antibiotic therapy. Chest radiography is necessary in all cases to outline basal lung changes (consolidation and effusion) and may show direct lung involvement in complicated cases. An elevated immobile diaphragm (radiological screening or ultrasonography) is often encountered, particularly in large abscesses. A plain film of the abdomen may demonstrate gas in the abscess cavity. Barium enema or colonoscopy may be indicated to exclude a colonic source of portal pyaemia.



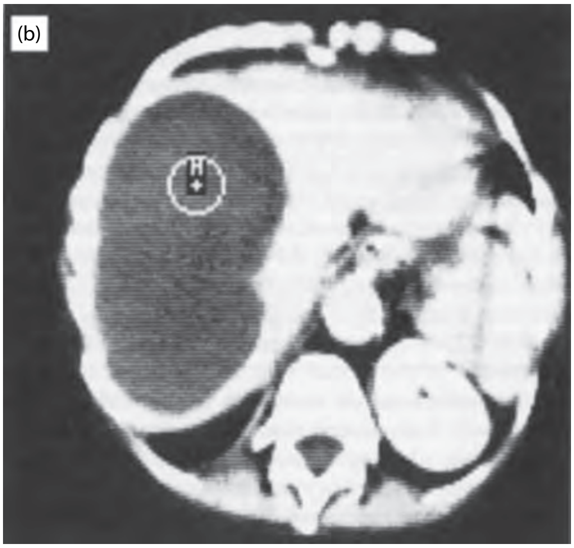


Figure 24.20 (a) An ultrasound of the liver on coronal section shows a cavity within the liver substance. (b) CT scan demonstrates that the abscess cavity occupies most of the right lobe of the liver. Subsequent drainage of the cavity resulted in complete resolution.

Complications

Recurrent bacteraemia is the most common complication of pyogenic abscesses. Extension and rupture of the abscess may occur in any direction. Peritoneal rupture results in widespread peritonitis or in the formation of a subphrenic collection. Extension through the diaphragm may lead to thoracic empyema or to a rupture into the bronchus with expectoration of large volumes of 'anchovy paste' pus from amoebic abscesses and bile-stained pus from cholangitic abscesses. Rarely, the abscess ruptures into the pericardium with high mortality.

Treatment

The key to successful management is drainage of the purulent collection combined with appropriate antibiotic therapy. Precise microbiological identification may result from blood cultures or from aspiration of the abscess cavity under ultrasound or CT guidance. In the event of a failure to isolate organisms, the choice of antibiotic should be based on the most likely aetiological factor, i.e. cholangitic and pylephlebitic abscesses are usually infected with *E. coli* and anaerobes and may be appropriately treated with cephalosporins, gentamicin and metronidazole. Microaerophilic *Streptococcus* infections are sensitive to penicillin. Although antibiotic therapy as the sole treatment is rarely successful, prolonged systemic antibiotic administration may be the only option for patients with multiple microabscesses.

The introduction and refinement of percutaneous drainage techniques have dramatically altered the management of patients with pyogenic hepatic abscesses. Percutaneous drainage has become the first-line therapeutic option in most centres for patients with single or multiple abscesses. Abscess communication with the intrahepatic biliary tree does not prevent pyogenic collections being successfully treated by percutaneous techniques, although the periods of drainage may be prolonged. Some groups have advocated the use of percutaneous aspiration combined with systemic antibiotics; however, in the only randomized trial comparing the two techniques, aspiration was successful in only 60% of patients whereas percutaneous catheter drainage was successful in 100% of patients. Regular irrigation of drainage catheters reduces the risk of catheter blockage with necrotic debris. Percutaneous catheter drainage is monitored by serial ultrasound examination to assess the degree of resolution. As the most frequent cause of hepatic abscess is biliary disease, effective decompression of the biliary tree is as important as abscess drainage where obstruction of the bile duct has contributed to the development of the hepatic abscess.

The factors indicating failure of initial non-operative management are:

- unresolving jaundice
- renal impairment secondary to clinical deterioration
- multiloculation of the abscess
- rupture on presentation
- biliary communication (Figure 24.21).

Deterioration in the general condition of the patient, repeated episodes of septicaemia or a failure of the abscess to decrease in size are indications for open surgical drainage. Posteriorly placed abscesses may be approached by the retroperitoneal route through the bed of the twelfth rib. Larger abscesses require abdominal drainage with wide-bore sump drains and special care being taken to avoid contamination of the peritoneal cavity. Where a pyogenic abscess is secondary to cholangitis, concomitant drainage of the common bile duct with a T-tube and removal of ductal stones may be necessary. Following successful drainage of the abscess, antibiotic administration should be continued for a prolonged period (3–6 weeks) to assist in the complete eradication of the infection.

Hepatic amoebiasis

This results from intestinal amoebiasis. The infection spreads to the liver via the portal vein from an ulcer in the bowel wall. Most amoebae lodge in the interlobular veins and degenerate. Some act by cytolysis to invade the portal tracts and lead to cell necrosis and a coalescence of infected triads to form a larger abscess cavity. Hepatic abscess is the most common extraintestinal manifestation of *Entamoeba histolytica*. There is a male preponderance and the mean age is 40 years. The presenting features include fever (77%), right upper quadrant abdominal pain (72%), cough (16%), chest pain (19%) and chest radiographic abnormalities (57%). The LFTs are often normal. The diagnosis is usually made by ultrasonography and CT where the boundaries of the abscess are usually poorly defined. Patients with amoebic hepatic abscess virtually always have a confirmatory serum antibody titre. Early abscesses are solid but pus appears later, characteristically resembling anchovy paste, though in one-third of patients the pus has the usual creamy appearance. Occasionally, amoebic abscesses become secondarily infected with pyogenic bacteria.

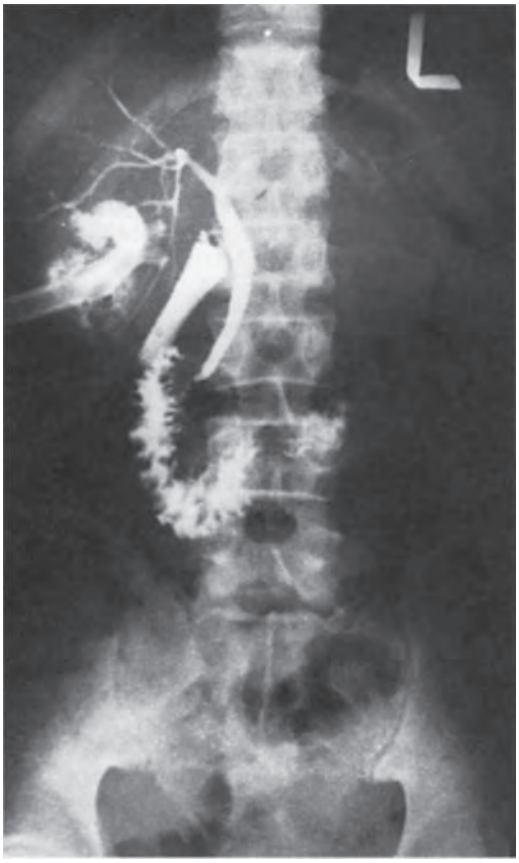


Figure 24.21 Right hepatic abscess communicating with the intrahepatic biliary tree.

In one large series from the USA, patients were divided into groups based on the presumed manner in which they had acquired ALA: (1) those born or raised in the USA, with a history of travel to an endemic area (Tr-ALA); (2) those from an endemic area, but living in the USA for less than 1 year (En-ALA); and (3) those neither from nor having travelled to an endemic area (N-ALA). In this series, there was a distinctive clinical pattern in patients from different epidemiological groups. Thus patients with Tr-ALA were a decade older than those from endemic areas and were more likely to be male, and to have an insidious onset. In addition patients in the Tr-ALA group were more likely to have hepatomegaly and large abscesses. Thirty per cent of patients had no associated travel history or endemic origin as risk factors (N-ALA). The majority of these had severe immunosuppression, such as infection with HIV, malnutrition or chronic infection.

The diagnosis of amoebic hepatic abscess requires confirmation by positive serological tests, demonstration of amoebic trophozoites in abscess fluid and a rapid response to antiamoebicides. Amoebic abscesses tend to occur in younger patients than in those with pyogenic liver abscess. In addition, they are much more likely to have abdominal pain and present with a history of symptoms of shorter duration. Both types, however, have the same high incidence of right lower lung abnormalities (50%). Amoebic abscesses are treated with metronidazole with or without chloroquin. Metronidazole therapy is invariably successful and has lowered the mortality of amoebic abscess, which is now negligible although recurrence after treatment can occur, albeit rarely. Aspiration of the amoebic abscess is not required routinely and is usually reserved for patients where serology is inconclusive or when there is no response within 2 days. Rupture of an amoebic abscess into the lung and bronchus can usually be treated successfully by antibiotics and postural drainage. When the intrapulmonary rupture is associated with cholangitis, bile duct drainage is essential. A persistent bronchopleural fistula may require formal thoracotomy, decortication of the lung and diaphragm with resection of severely damaged pulmonary tissue and diaphragmatic repair. Rupture into the pericardium requires early aspiration of the exudate and occasionally transpleural drainage. In a meta-analysis of 3081 patients with ALA, the mortality rate was 4%, compared with a mortality rate of 46% in patients with pyogenic liver abscess.

Subphrenic extrahepatic abscess

The distribution of intra-abdominal abscesses is directly related to the precipitating lesion and to the potential peritoneal spaces. Abscesses are most common in the right and left lower quadrants as a consequence of appendicitis, diverticulitis and pelvic sepsis. Abscesses around the liver make up the next most common group of intra-abdominal septic collections. Six spaces around the liver are described. Superiorly, the falciform ligament divides the left and right subphrenic spaces, the latter being divided into anterior and posterior spaces by the triangular ligament. In the infrahepatic region, there are also three spaces – the two on the left side including the lesser sac of the peritoneal cavity and the space immediately below the lateral lobe of the left hepatic lobe (Figure 24.22).

Subphrenic collections may contain a mixture of pus and gastrointestinal secretions and may be large. Although any intra-abdominal sepsis may precipitate these collections, most abscesses follow operative intervention with postoperative leaks, spontaneous perforation of hollow organs or accidental abdominal trauma. In the last situation, extrahepatic collections of blood or serous fluid become encysted and infected (Figure 24.23). Primary causes include pancreatitis, cholecystitis, perforated peptic ulcer, perforated diverticular disease and perforated acute appendicitis.

Clinically, suspicion of an extrahepatic collection is aroused by fever with occasional episodes of septicaemia. Non-specific abdominal symptoms and general ill health are common, although

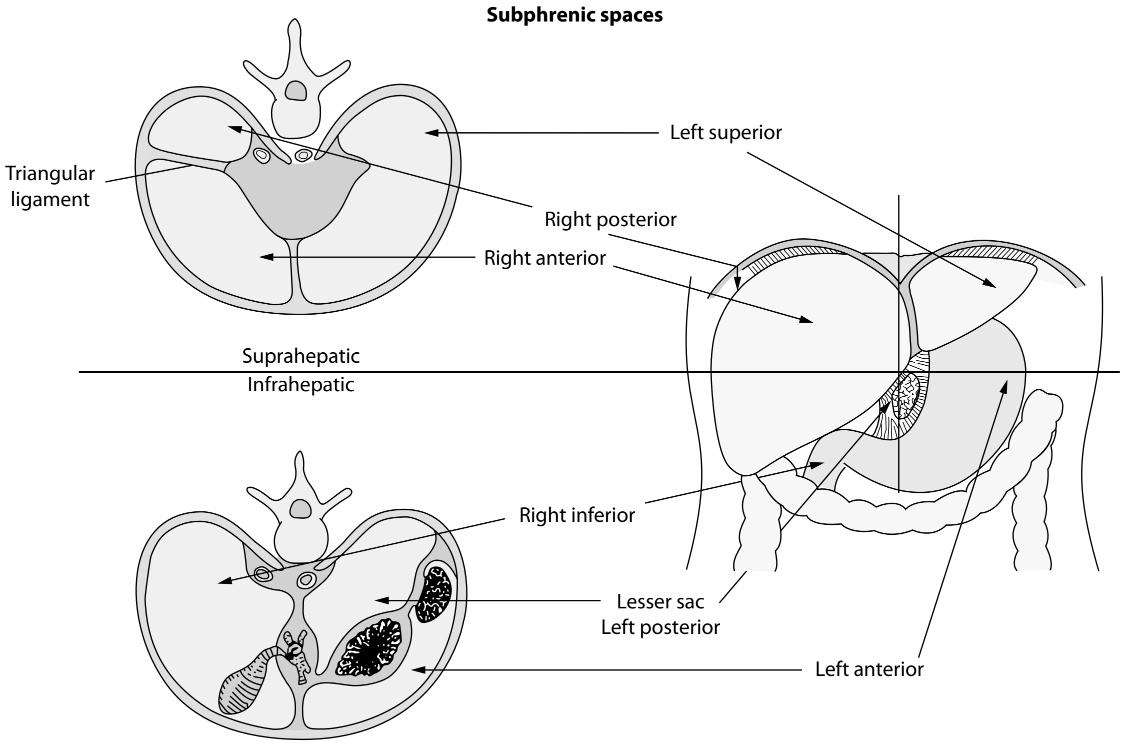


Figure 24.22 The subphrenic spaces may be considered to be supra- and infrahepatic.

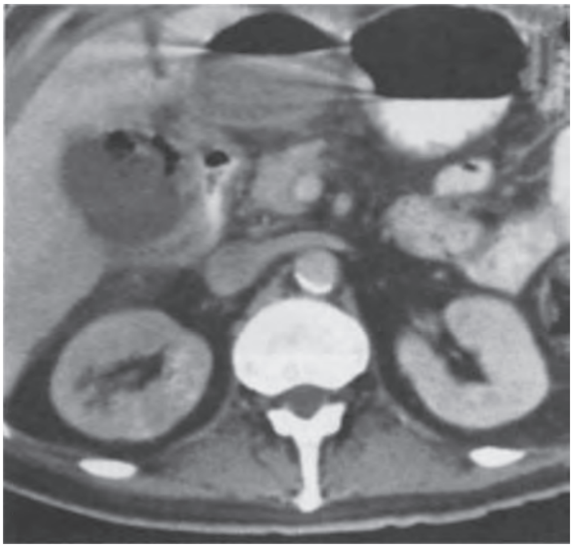


Figure 24.23 CT scan of an infrahepatic abscess with fluid level lying between the gallbladder and gastric antrum. The abscess resulted from the infected gallbladder which contains stones seen on other sections of the CT examination.

in some patients attention is drawn to respiratory symptoms resulting from inflammatory changes induced in the lung by a subdiaphragmatic collection. When an abscess is subhepatic, tenderness is elicited and a mass may be palpated or percussed.

The confirmation of an abscess may be difficult and probably many minor collections ultimately resolve and remain unproven. In right posterior subphrenic space collections there is typically elevation of the diaphragm on chest radiography, with decreased mobility of the diaphragm on screening. Inflammation of the diaphragm leads to pneumonic changes in the overlying lung and a pleural effusion. However, diagnosis and location of subphrenic abscesses is nowadays made by ultrasound and CT scanning. When the patients are stable, CT is the imaging modality of choice for most intra-abdominal processes. CT can detect small volumes of fluid, areas of inflammation and other gastrointestinal tract pathology with high sensitivity.

Management

Complicated intra-abdominal infections are an important cause of morbidity and are frequently associated with poor prognosis, particularly in higher risk patients. Mortality rates associated with complicated peritonitis associated with severe sepsis have a reported mortality of approximately 30%. Factors consistently associated with poor outcome in patients with intra-abdominal infections include increased illness severity, failed source control, inadequate/inappropriate antimicrobial therapy and nosocomial pathogens. The cornerstones in the management of complicated intra-abdominal infections are both source control and antibiotic therapy.

Minor collections may resolve with antibiotics and the frequency of dense adhesions between the liver and diaphragm at elective second laparotomies suggests this is a relatively common event. However, once pus is identified in the subphrenic region, drainage of the abscess is mandatory. Timing and adequacy of source control are the most important issues in the management of intra-abdominal infections, because inadequate or late control may have a negative effect on outcome. With accurate localization using ultrasound or CT scanning, many collections can be drained percutaneously with a catheter left in place. Numerous studies in the surgical and radiological literature have documented the effectiveness of percutaneous drainage to the extent that it is now the standard treatment. In selected cases percutaneous drainage has been shown to be associated with decreased incidence of complications, decreased rates of inadequate drainage and reduced duration of drainage compared with open drainage.

In most instances percutaneous drainage suffices but some require open surgical drainage. These include multilocular collections and the presence of thick pus with necrotic slough. Most subhepatic abscesses may be drained surgically by an anterior approach, as may the right anterior subphrenic collection. Occasionally it is possible to enter the abscess without contamination of the peritoneal cavity. Posterior subphrenic collections are most appropriately drained by the classic extraserous approach through the bed of the twelfth rib. In either approach, the abscess cavity is thoroughly explored and all loculi are broken down. Wide-calibre drains are inserted to the furthermost or most dependent point of the cavity. Gradual withdrawal of the drains is indicated, sometimes with regular saline irrigation. Further cultures (from drain effluent) are often needed especially if resolution is delayed.

Abscesses occurring as a direct consequence of an uncomplicated surgical operation are largely preventable. Prior to abdominal closure, a saline lavage of the peritoneal cavity with special attention to the subphrenic spaces and pelvis will reduce the incidence of postoperative abscesses by removing blood clots and other secretions that may get infected at a later stage.

Liver cysts

These include simple hepatic cysts, choledochal cysts, cystadenoma of the liver, Caroli disease and parasitic cysts. Choledochal cysts are covered elsewhere (Chapter 25).

Non-parasitic hepatic cysts

Within the non-parasitic group there are a variety of clinical conditions that reflect underlying developmental defects of the liver parenchyma or bile ducts. Some cystic lesions follow trauma in which a central rupture has resulted in a collection of bile and serum and these cysts have no epithelial lining. Others are clearly similar to dermoid cysts found in other sites. Most of the remainder are lined by cuboidal or columnar epithelium, contain serous fluid and do not communicate with the biliary tract. These are nowadays referred to as simple cysts and are multiple in 50% of cases. The cysts can grow to a large size and in so doing cause pressure atrophy of the surrounding hepatic parenchyma. They are generally regarded as developmental abnormalities from aberrant bile ducts. The ultrasound incidence of asymptomatic cysts is 1%, but symptomatic cysts are much rarer. Symptomatic cysts are much commoner in females (9:1) and huge cysts are almost exclusively found in women above the age of 50 years (Figure 24.24).

The ciliated hepatic foregut cyst (CHFC) is a rare small to medium size solitary cystic lesion of the liver that is commoner in males (as distinct from the solitary hepatic cysts) and occurs over a wide age range (35–75 years). CHFC is most commonly located in the medial segment of the left hepatic lobe as distinct from the more common solitary hepatic cysts. CHFC is usually found incidentally on radiological imaging or during surgical exploration, although some may present with abdominal pain, and in a collective review of 52 cases one patient presented with portal vein compression. Histologically, the lining of the columnar epithelium is composed of ciliated cells, mucin-secreting goblet cells and endocrine cells positive for chromogranin, synaptophysin, bombesin and calcitonin (similar to respiratory epithelium). The lesion is thought to be a developmental ventral foregut abnormality arising from a bronchiolar bud of the tracheobronchial diverticulum.

Clinical features and treatment

Most simple hepatic cysts are asymptomatic and only become apparent when the cysts reach sufficient size to exert pressure on adjacent viscera, producing non-specific symptoms of vomiting, upper abdominal pain and occasionally diarrhoea. Clinical examination reveals a non-tender smooth mass in the liver. Jaundice is very unusual and LFTs are usually normal. Plain film of the abdomen may show displacement of the colon or stomach but the diagnosis is best confirmed by ultrasonography. Other investigations including CT are not usually necessary. Complications are uncommon and include intracystic bleeding that causes sudden severe pain and increase in size, fistulation with the intrahepatic biliary tract or duodenum, bacterial infection, compression of the bile duct with obstructive jaundice and compression of the vena cava or portal vein. Differentiation is from parasitic cysts and from adult polycystic disease of the kidney where multiple serous hepatic cysts are often present and may indeed replace a substantial part of the hepatic parenchyma.

Only symptomatic cysts require treatment. This consists of fenestration (deroofing), which is nowadays carried out laparoscopically. Percutaneous aspiration of large cysts with introduction of sclerosing agents is followed by a high recurrence rate. In patients with adult polycystic disease of the kidney and liver cysts causing discomfort, multiple fenestration or hepatic resection is followed by improvement but recurrence is inevitable although symptomatic improvement may last for up to 2 years.

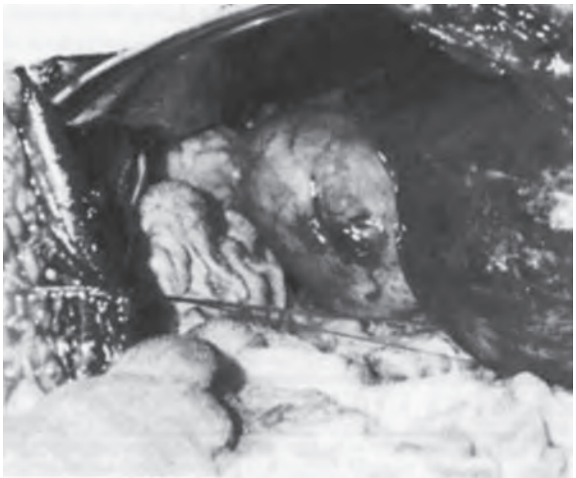


Figure 24.24 This operative photograph shows the undersurface of the liver in which the gallbladder is seen as a whiter structure and lies adjacent to a large lymphogenous cyst of the liver.

Hepatic cystadenoma

This is rare and affects predominantly females. It forms a large multiloculated cyst filled with mucinous fluid and lined by cuboidal epithelium on a basement membrane and thick compact cellular stroma containing foamy macrophages. In places the lining epithelium forms polypoid projections. Although benign, the lesion is liable to complications, notably cholestasis due to compression of the bile duct, intracystic bleeding, infection, rupture and malignant degeneration to cystadenocarcinoma. Hepatic cystadenoma must be excised completely even when asymptomatic. Fenestration/partial resection is inevitably followed by recurrence and increases the risk of cystadenocarcinoma (Figure 24.25).

Caroli syndrome

This is not a single entity and covers a spectrum of disorders characterized by congenital multifocal dilatations of the segmental bile ducts. In 50% of cases Caroli syndrome is associated with congenital hepatic fibrosis, itself an inherited malformation (autosomal recessive). The clinical picture of Caroli syndrome is dominated by recurrent episodes of bacterial cholangitis (Chapter 25).

Hydatid cysts of the liver

Hydatid disease is a parasitic infestation by a tapeworm of the genus *Echinococcus*. It is endemic in certain countries, particularly the southern half of South America, Australasia, New Zealand, France and certain areas of the USA and the UK. Humans are secondary, incidental hosts while dogs and their related species are the definitive host of the parasite. The worm can be found also in intermediate hosts like sheep, goats or swine. Humans become infected by ingesting vegetables or water contaminated by dog excrement or more directly by handling the parasite-

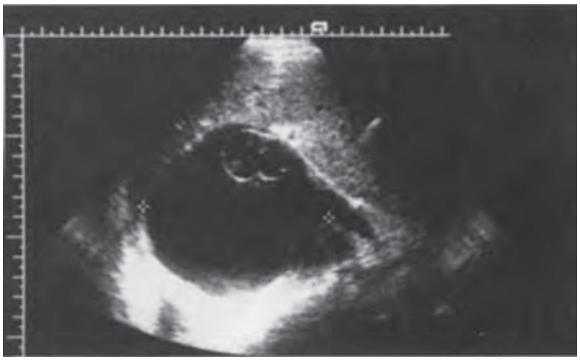


Figure 24.25 Ultrasound scan of hepatic cystadenoma. The lesion recurred following initial deroofing and eventually needed a right hepatectomy.

infested dogs as pets. After ingestion, the egg is destroyed by gastric acid and the embryos hatch within the duodenum and then migrate through the gut wall into the mesenteric circulation to lodge within the liver. Echinococcosis can involve any organs but the liver is the most common site followed by the lungs, which together account for 90% of all cases. The embryo becomes a small vesicle with an inner germinal epithelium that produces secondary or daughter brood cysts containing scolices and hooklets. Some cysts never produce brood capsules or become sterilized by secondary infection or calcification. Hydatid cysts caused by *Echinococcus granulosus* are unilocular as distinct from the multilocular alveolar type due to *Echinococcus multilocularis.* This forms a spongy collection of cysts and carries a poor prognosis.

Clinical features

Since the growth of the parasite is slow, many years elapse before the cyst reaches significant size (Figure 24.26). Palpable cysts are therefore rare in children. At all ages, pain, jaundice and ascites are uncommon and, in most patients, general health is good. On physical examination, an anteriorly located cyst presents as a smooth rounded tense mass. Secondary infection results in tender hepatomegaly, rigors and pyrexia associated with a deep-seated continuous pain. Further clinical features are the result of cyst complications. Intrabiliary rupture may give biliary colic and usually causes jaundice and fever. Intraperitoneal rupture produces severe pain and shock classically associated with pruritus and urticaria. Some of the implanted brood cysts induce a profound fibrous reaction that sterilizes the infection, but other cysts reappear in various parts of the peritoneal cavity years later. Intrathoracic rupture may be preceded by symptoms of diaphragmatic irritation, and rupture into bronchus leads to a partly blood-stained sputum which frequently becomes bile stained. Hydatid allergy is manifested by urticaria or, very rarely, anaphylactic shock.

Investigations

An unruptured cyst may show on plain radiograph as a calcified reticulated shadow or by displacement of the diaphragm and stomach if not calcified. Following intrabiliary rupture, gas may enter the cyst leading to partial collapse of the cyst wall (Camellotte sign). Ultrasonography reveals an echogenic cyst.

Although the cyst is isolated from the liver by an adventitial layer, there is absorption of parasitic products, which acts as an antigenic stimulus. This is reflected in eosinophilia in 25% of patients, and a positive complement fixation test which is accurate in 93% of patients. The Casoni test has been largely abandoned. Some cysts never leak and tests are never positive in these patients.

Treatment

The treatment of hydatid cysts of the liver is surgical or radiological. There is no guaranteed response to drug treatment with mebendazole/albendazole and the cyst is a potential source for serious complications. Surgical treatment involves removing the cyst without contaminating the patient. Where there are

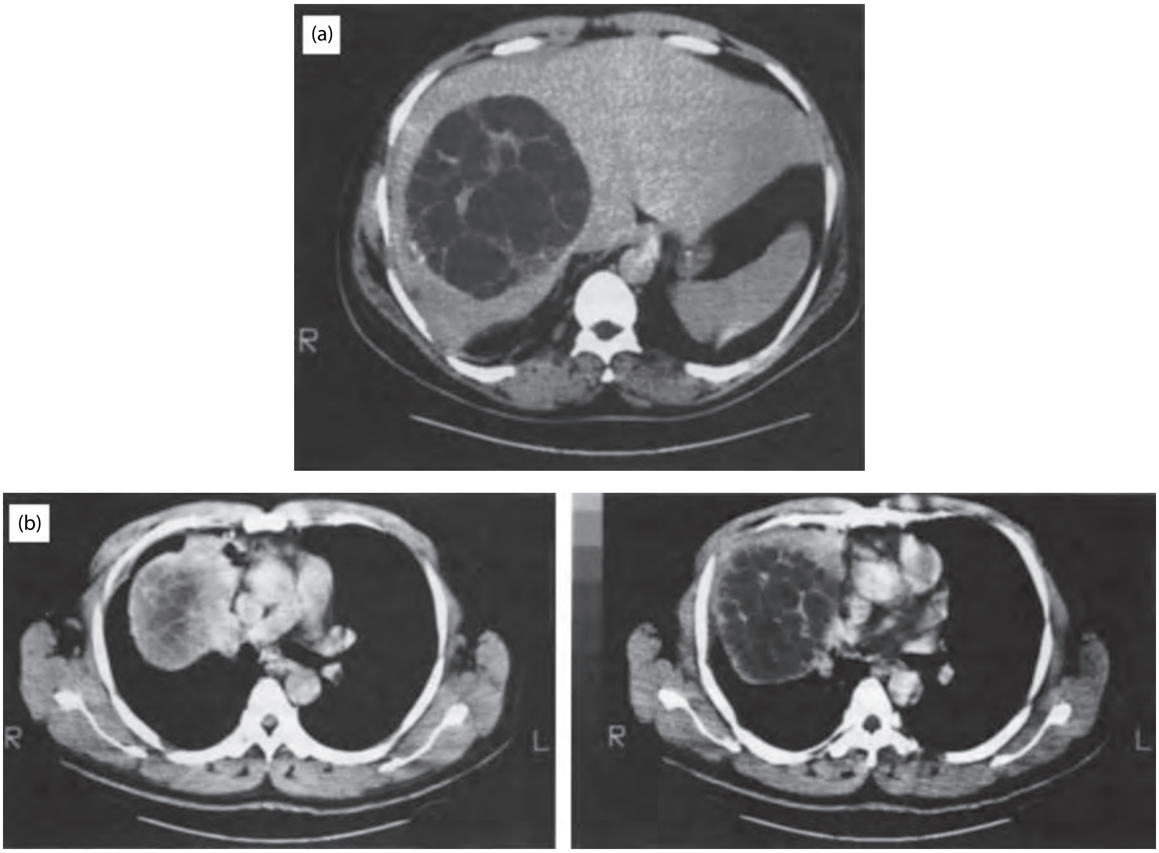


Figure 24.26 CT scan showing multiseptate hydatid cyst (a) in the right lobe and (b) projecting above the right liver surface in the subphrenic space.

multiple cysts, several procedures may be necessary. Large cysts found on the anteroinferior and posteroinferior aspects of the liver are approached abdominally. Cysts in the dome may be reached by the posterior extraserous approach or transpleurally through the bed of the ninth rib.

The initial stage involves protection of the operative field against live cysts using multiple coloured towels soaked in hypertonic saline which isolate the main cyst from the exposed serous cavity. Since hydatid fluid is under high pressure, the cyst is decompressed by aspiration as completely as possible, although daughter cysts tend to block the needle frequently. The main cyst is injected through the same needle with 20% hypertonic saline and left for 5 minutes, after which the main cyst is opened and all daughter cysts are removed. In large cysts it is not feasible to remove the cyst wall and the cavity is drained for a few days and partially occluded by an omental plug which is sutured to the rim of the cyst.

Marsupialization of large cysts may be indicated when secondary infection has occurred but prolonged purulent drainage results and a secondary omentoplasty may be necessary. Cysts with extensive calcification are usually sterile and best left alone. Jaundice after intrabiliary rupture requires choledochotomy and clearance of cysts followed by T-tube drainage. Peritoneal rupture is managed by laparotomy and careful toilet followed by lavage. Providing that rupture has not occurred and careful surgical techniques are applied, the prognosis is excellent.

There have been good reported results of percutaneous ultrasound-guided treatment of hepatic hydatid cysts. The treatment involves puncture, aspiration, injection (of scolicidal agents), and reaspiration (PAIR). In patients with cysts larger than 6 cm in diameter, PAIR is followed by percutaneous drainage (PAIR-PD). In one recent large reported series no recurrence of hydatid disease after PAIR or PAIR-PD was observed during a follow-up period of 72 months (mean 26 ± 27 months).

Medical treatment with mebendazole and albendazole is unreliable as penetration of the drug into the cyst cavity is uncertain. Thus although many cysts decrease in size, not all scolices are killed. Nowadays medical treatment with these drugs is started before surgery or radiological intervention (PAIR).

Liver tumours

Benign solid tumours

Benign tumours of the liver are commonly asymptomatic and discovered accidentally during investigation usually by ultrasound, and during laparotomy for other conditions. More rarely, they reach sufficient size to cause symptoms or complications that can be quite dramatic. Although the classification of these tumours is not sharply defined several types are identified:

- haemangioma
- infantile haemangioendothelioma
- neoplastic angioendotheliomatosis
- hamartoma
- FNH
- adenoma
- cholangioma
- biliary cystadenoma.

Haemangiomas

Haemangiomas are the commonest benign tumour of the liver most commonly found in adults between the ages of 30 and 70 years, but only rarely produce symptoms. They are frequently situated just beneath the liver capsule and are normally of the cavernous type. Histologically the lesion is composed of blood-filled endothelium-lined spaces separated by a variable degree of fibrous tissue and inflammatory changes, both of which result from episodes of spontaneous thrombosis.

Infantile haemangioendothelioma is the most common mesenchymal tumour in this age group and the majority present in the first few months of life. The lesion may be solitary or multicentric and often extends over a wide area. Histologically it consists of numerous arteriovenous channels, endothelium-lined vascular lakes in a fibrous stroma containing bile ducts. The tumour is locally aggressive and presents with hepatomegaly, coagulopathy (due to thrombosis in the tumour), extramedullary haemopoiesis, cardiac failure (commonest presentation), bleeding and liver failure. Most deaths are due to cardiac failure. A related but histologically different condition is *neoplastic angioendotheliomatosis.* Although essentially benign these diffuse subcapsular haemorrhagic lesions in children have a variable course. They may mature to a cavernous haemangioma or regress spontaneously, but rarely they may undergo sarcomatous change. In children the condition may present as high-output cardiac failure and is often associated with cutaneous haemangiomas (85%). Other complications are attributed to recurrent thrombotic episodes in the tumour. These include microangiopathic anaemia, thrombocytopenia and hypofibrinogenaemia.

In adults, some of these cavernous haemangiomas, having grown to significant size (>4.0 cm), will eventually produce pain or dyspepsia and may develop a palpable abdominal mass. A bruit is heard in about 15% of patients. The pain is usually dull but acute posterior pleuritic pain lasting a few weeks may develop and has been attributed to episodes of intratumoral thrombosis and inflammation. In general symptomatic cavernous haemangiomas are more common in females and have been documented to enlarge during pregnancy. Intratumoral thrombosis may be precipitated by oral contraception.

Rupture of cavernous haemangioma is rare. The reported incidence of this complication varies from 0 to 10%. Nonetheless, it is a serious complication in view of the major intra-abdominal haemorrhage with shock and collapse. Cavernous haemangiomas have characteristic appearances on the various imaging modalities:

- hyperechoic lesion on ultrasound
- hypodense on CT without contrast, halo edge enhancement on high-dose contrast helical CT which then fills the centre of the lesion
- dark on T_1 and intensely white on T_2 weighted MRI images.

The need for arterioportography (Figure 24.27) seldom arises nowadays. A biopsy is not indicated.

The majority of these tumours do not require any treatment except for follow-up liver ultrasound examination. The preferred treatment for clinically significant/symptomatic haemangiomas

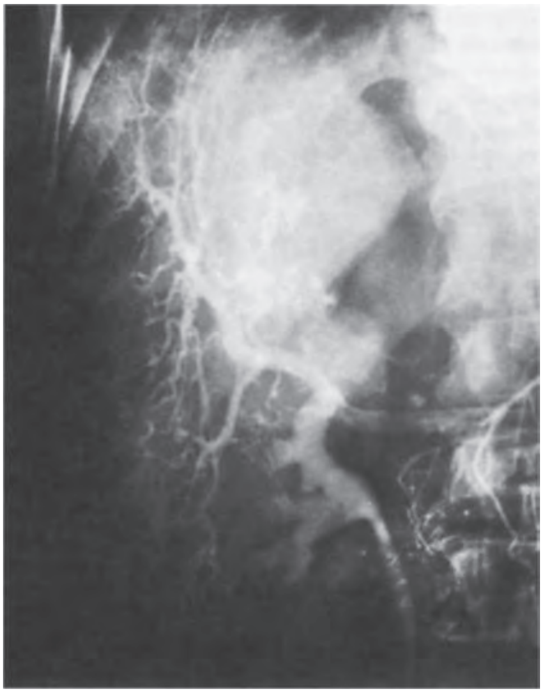


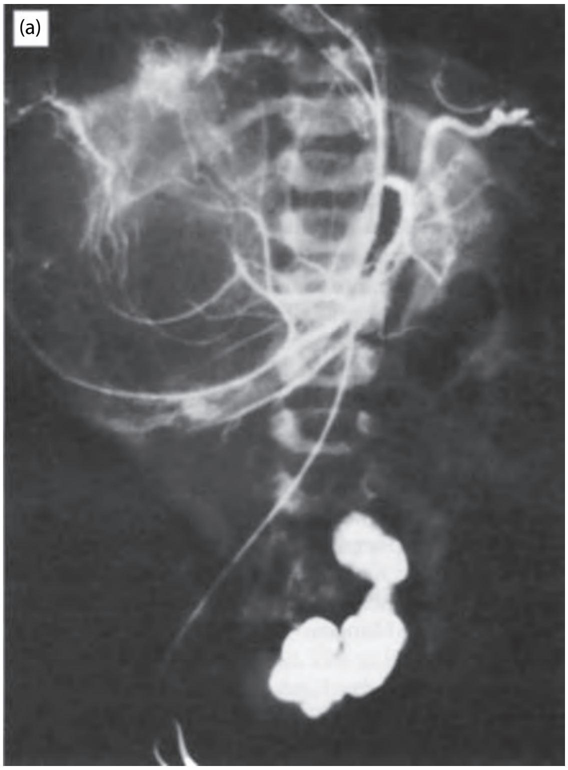
Figure 24.27 This selective arteriogram shows an apparent tumour circulation arising from the left hepatic artery. Subsequent resection of a tumour in the left lobe showed this to be a benign cavernous haemangioma.

is segmental/wedge excision where possible, with lobectomy being reserved for large/multiple lesions confined to one lobe. In such cases the residual liver may contain further haemangiomas. Supervoltage radiotherapy is used for symptomatic lesions in adults who are unfit for surgery.

Children presenting in the first year of life have a poorer prognosis. Hepatic angiography in this age group may demonstrate a major feeding vessel from the hepatic artery, when ligation of this vessel or the main hepatic artery may reduce the blood flow through the lesion. With more diffuse lesions, radiotherapy and steroids result in significant shrinkage and a reduction in the volume of arteriovenous shunting. It is nowadays the preferred management in this age group. After the first year, the angioendotheliomatosis may regress, as do the cutaneous lesions. Resection of vascular tumours is no longer advocated in children, but liver transplantation is the correct treatment for infantile haemangioendothelioma that fails to respond to radiotherapy.

Mesenchymal hamartomas

Hamartomas are tumour-like malformations of congenital origin and consist of normal tissues in a disorderly arrangement. The lesions vary from minute nodules to large solid tumours and may be single or multiple. The tumours are not encapsulated, the fibrous periphery is a pseudocapsule of compressed parenchyma (Figure 24.28) and the histological picture demonstrates irregular distorted hepatic plates, vascular channels, bile ducts, cysts and extensive fibrosis. Children are commonly affected and present with an expanding abdomen and a large palpable mass. Large tumours may displace the stomach and produce vomiting and elevate the diaphragm compressing the right lung. As the tumour expands, inferior vena caval compression occurs.



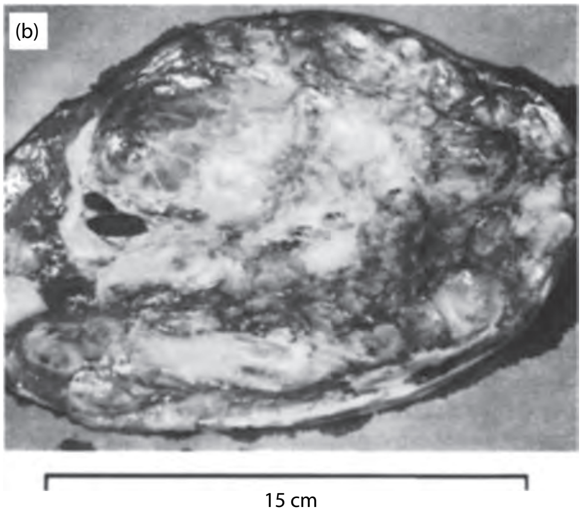


Figure 24.28 (a) A 3-year-old child presented with an expanding liver tumour. Selective arteriogram showed the hepatic vessels to be stretched around the tumour and there is little evidence of tumour circulation. (b) The excised lesion proved to have a distinct capsule and to be multilocular. Histology showed this to be a hamartoma of the liver.

Mature hepatic teratomas have been reported. All have mesodermal, endodermal and ectodermal components (tridermal). They may be associated with chromosomal abnormalities, e.g. trisomy 13. Hamartomas show as filling defects in angio-CT or scintiscans as they are relatively avascular tumours. Large tumours should be removed, usually by formal lobectomy. Occasionally, large tumours may show evidence of sarcomatous change and have a poor prognosis.

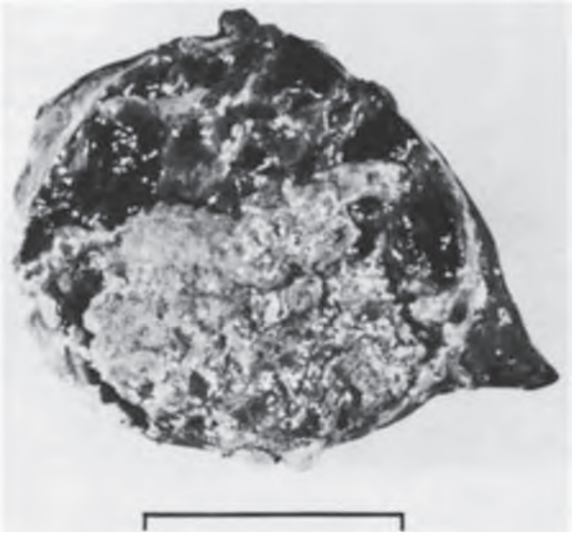


Figure 24.29 This operative specimen of the right lobe of the liver has been cut to show the necrotic contents of a hepatic adenoma into which a major haemorrhage has occurred. The patient presented with severe upper abdominal pain and evidence of massive blood loss.

Benign lesions have an excellent prognosis and when small and multiple should be kept under observation.

Liver cell adenomas

Adenomas are variable in size (4–30 cm). They are rare and tend to occur most commonly in children and postmenopausal women. There is an established increased incidence related to use of the contraceptive pill. The risk rises after 4 years of pill usage, particularly in women over 30 years on pills of high oestrogen content. Liver cell adenoma can also occur in patients with type I glycogen storage disease and galactosaemia. Up to one-third of adenomas are multiple.

Macroscopically liver cell adenoma forms a pale soft smooth lesion without a fibrous capsule. Histologically the tumour consists of sheets of hepatocytes containing glycogen, venous lakes and necrotic ghost cells. The differentiation between liver cell adenoma and well-differentiated hepatocellular (lamellar) carcinoma may be difficult and the two may coexist. Furthermore HCC may develop in patients with biopsy-confirmed liver cell adenoma. In the absence of a well-defined central scar, the differentiation between an adenoma and FNH can also be difficult. In this case technetium-99 sulphur colloid scintigraphy can be useful. Liver cell adenomas do not take up the Tc-labelled tracer and have a typical photopenic appearance, explained by their supposed lack of reticuloendothelial cells. Altered blood flow (secondary to intratumoral infarction and haemorrhage) or decreased activity of Kupffer cells may also help explain the appearance of adenomas on scintigraphy. More recently the sensitivity of a Tc-colloid scan can be increased with the use of simultaneous single photon emission computerized tomography.

The symptoms of liver cell adenoma include upper abdominal pain and mass. The lesion carries an ever-present risk of rupture and haemorrhage, a complication that occurs in 30% of cases (Figure 24.29). A specific diagnosis of liver cell adenoma cannot be made by CT, ultrasound or MRI. Although they have a distinctive arteriographic appearance (enlarged tortuous vessels, peripheral and central feeding), this is also encountered in HCC. Nowadays diagnosis is established by ultrasound or laparoscopically guided biopsy. The true incidence of malignant degeneration of hepatic adenoma is not known. It has however been widely reported that malignant transformation can occur after an interval period of 2–5 years and that foci of HCC can often be found in surgical specimens of liver adenomas (10–35%), but its reported frequency seems to be around 5% when observed prospectively. It appears to be 10 times more common in males than females, which would suggest a different carcinological pathway, and seems to be a very unusual complication in females with lesions smaller than 5 cm.

Treatment

If the diagnosis is certain and the lesion is asymptomatic, it may be monitored by ultrasound and AFP for 6–12 months for at least 10 years. In addition, oral contraception should be stopped. Adenomas that diminish in size can be managed expectantly, whereas those that do not regress on follow-up should be excised. If the patient becomes pregnant, the risk of rupture may be increased.

Ruptured and bleeding tumours are excised as emergency procedures and, since many of these tumours are near the surface, local excision with a thin rim of parenchyma is sufficient, although whenever possible the resection should be anatomical. Intracapsular haemorrhage may produce an extremely large tumour necessitating formal lobectomy.

Focal nodular hyperplasia

FNH (Figure 24.12b) is nowadays regarded as a separate entity from liver cell adenoma on aetiological and pathological grounds. FNH is best considered as a response to parenchymal injury or to an anomalous arterial supply to a local area of liver tissue. The lesion is usually small and may be multiple (20%), although large lesions presenting as an abdominal mass may occur especially in pregnant females and in children of both sexes. There is no aetiological association with oral contraception and the incidence of FNH has not increased since the introduction of the contraceptive pill. Macroscopically FNH consists of a firm mass, the cut surface of which reveals a central scar with radiating fibrous septa. The microscopic findings are akin to cirrhosis with regenerating nodules and fibrosis. The vast majority of FNH cases are discovered accidentally during surgery or investigation. FNH is not premalignant and the natural history is such that it may be observed without serious risk.

Adenomatous hyperplasia

This ill-defined pathological entity refers to sizeable nodules that develop in chronic liver disease. In cirrhosis, adenomatous hyperplasia forms part of the spectrum of morbid pathological change which includes small benign regenerative nodules, large size regenerative nodules (adenomatous hyperplasia), atypical adenomatous hyperplasia (dysplastic borderline lesions, considered as low-grade HCCs by some pathologists) and frank HCC. There is some debate as to the malignant potential of adenomatous hyperplasia as some reports have indicated that cancer can develop from the intervening hepatic parenchyma, whereas others consider the transition from regenerative nodules to adenomatous hyperplasia, dysplasia and carcinoma to be the norm.

Cholangioma (Von Meyenburg complexes)

These are derived from bile duct epithelium and form small greyish white nodules consisting of mature bile ducts and fibrous tissue with a cystic component. The condition is entirely asymptomatic and requires no treatment. Most are diagnosed on biopsy for nodules found accidentally during surgery. Von Meyenburg complexes (VMCs) do carry a small risk of malignant transformation (only 11 cases of neoplastic transformation of VMCs have been reported). The malignant tumour that may develop from VMCs is cholangiocarcinoma usually in patients over 60 years of age.

Primary malignant tumours of the liver

The primary malignant tumours of the liver are dominated by HCC. By comparison, other primary tumours are rare or very rare. They include hepatic cholangiocarcinoma, angiosarcoma and epitheliid haemangioendothelioma.

Hepatocellular carcinoma, hepatoma

Primary HCC is the commonest malignant tumour worldwide and its geographical distribution parallels closely the incidence of hepatitis B and C viral infection. Although the highest prevalence is encountered in the Far East and sub-Saharan Africa, it is also common in selected populations within the USA and Europe. The disease usually develops on a background of cirrhosis (cirrhomimetic) but can originate in normal or non-cirrhotic hepatic parenchyma (non-cirrhomimetic). In the South African black population, where the incidence is high (HCC accounts for 80% of all cancers in the Bantu), 37% of primary hepatic carcinoma occur in patients without cirrhosis. This is a much higher incidence of non-cirrhomimetic disease than anywhere else in the world, including Japan (11%). In general, there are no marked differences between patients with and without cirrhosis in the symptomology, size of tumours at presentation, serum markers and HBV status, although a lower incidence of HBsAg is reported in some countries in non-cirrhomimetic cases. It is important to stress that a significant cohort of non-cirrhotic patients have histological evidence in the portal tracts of ongoing chronic liver disease.

The disease is more common in males and, although slow growing, it is associated with a poor prognosis and a frequently fatal outcome with a median survival of 4 months in patients with symptomatic disease. The outcome is not influenced by racial origin or the presence of chronic liver disease and is largely determined by the stage of the disease.

Aetiological factors

The exact pathogenesis remains unknown but the development of the disease seems to occur in stages with initiation by a genotoxic event followed by transformation due to the action of co-carcinogens. There is no evidence for a genetic predisposition, and the high incidence in certain ethnic groups and regions is explained by the correspondingly high prevalence of chronic viral liver disease in these populations.

There is undoubtedly an association between the development of HCC and the presence of chronic (necroinflammatory) liver disease with or without a background of HBV/HCV disease. One hypothesis, suggested by Dunsford *et al.*, is that, irrespective of the aetiology, chronic liver cell injury induces a series of events (cell death, regeneration, cellular metabolic dysfunction and release of inflammatory mediators) which collectively increase the risk of transforming mutations. Chronic liver disorders which are particularly prone to the development of HCC are haemochromatosis, ALD and hereditary tyrosinaemia.

The importance of HBV infection in the development of HCC is undoubted. The risk is increased 94-fold in HBsAg-positive males. In the Far East, 90% of patients who develop HCC are HBsAg positive, although the corresponding figure in the West is substantially lower (10–30%). It is now known that hepatitis B viral DNA and RNA are present in tissue extracts of the tumour and the viral DNA is commonly integrated in the host genome. However, since the non-neoplastic parenchyma adjacent to the tumour also contains viral DNA, integration alone does not account for the carcinogenesis, and the molecular mechanisms by which the hepatitis B viral DNA induces mutation remain obscure. It has been suggested that, subsequent to integration, promoting cofactors are important and these include specific carcinogens and the known risk factors, i.e. male sex, alcohol, iron overload and cigarette smoking. Again the molecular mechanisms involved in the action of these cofactors are not known. The other hypothesis concerning the development of HCC in HBV disease implicates the insertion of the viral DNA close to a cellular proto-oncogene, which is then stimulated to produce cellular transformation. Alternatively, the integration of the viral DNA is followed by chromosomal rearrangement leading to carcinogenesis.

Since the identification of hepatitis C, there have been several reports, particularly from southern Europe (Italy and Spain) and the USA, that have shown a significantly high incidence of antibodies to HCV in patients with HCC. The mechanism is obscure since HCV is not a retrovirus. It is assumed that its oncogenic action is the result of a chronic necroinflammatory state along the Dunsford hypothesis. Many, but not all, of these patients have had alcoholic or HBsAg chronic liver disease. It is now generally accepted that hepatitis C disease predisposes to the development of HCC.

Coinfection with HBV and HCV is not uncommon and results in more severe liver disease and with a significant increased risk for the development of HCC. In one recent study from Japan, 59% of patients with HCV disease and who developed HCC had evidence of past HBV infection and were anti-HBc positive. The mechanism for this synergy between coinfection with HBV and HCV is not known but it has been suggested that the HBV-encoded X protein that regulates both cell proliferation and apoptosis on a background of increased hepatocyte turnover present in patients with chronic HCV infection may be involved.

Of the mycotoxins, the most important is aflatoxin B, produced by a fungus that contaminates grain and nuts, particularly in West Africa. Recently, aflatoxin B has been detected histochemically in patients with HCC. The aetiological role of hepatic helminthosis remains obscure and the only firm association is between infestation with *Clonorchis sinensis* and the development of cholangiocarcinoma as distinct from HCC.

Pathological features of hepatocellular carcinoma

From a macroscopic viewpoint HCCs are described as follows:

- *Expanding* – sharp demarcation between tumour mass and the compressed (sometimes atrophied) surrounding parenchyma. Some have a definite capsule. Further classified into single nodular or multinodular.
- *Pedunculated* – predominantly extrahepatic growth. This has two subtypes: type I intrahepatic origin, or type II extrahepatic tumour mass that becomes nourished by a branch of the right hepatic artery (not true HCC).
- *Spreading* – lack of demarcation between tumour and surrounding hepatic parenchyma. May involve the whole liver as nodules (usually in cirrhotic livers) or be diffusely infiltrating (non-cirrhotic livers), sometimes referred to as Engel's massive form.
- *Multifocal* – several unconnected small tumours of similar size that appear to have arisen synchronously rather than as metastatic foci from an initial single tumour.
- *Indeterminate* – no distinct pattern, different features in different parts.

Another perhaps more important distinction is *small HCC*. According to the *Japanese General Rules for Clinical and Pathological Study of Primary Liver Cancer*, this is an HCC of a diameter that is less than 2.0 cm. However, during the last 10 years, this limit has been extended by surgeons to tumours smaller than 3.0–5.0 cm. In any event small HCCs are nodular and well differentiated. A capsule is not identified when the lesion is smaller than 2.0 cm, but larger tumours possess a definite capsule. Histologically, small HCCs are difficult to diagnose and in particular to differentiate from large regenerative nodules. A suggestive histological picture with a rising AFP is the basis for establishing the diagnosis. Small HCCs are slow growing and have an excellent prognosis with resection or *in situ* ablation. Recurrence and prognosis are influenced by size, extracapsular growth and involvement of the portal venules.

Some large tumours may have necrotic centres. On cut section there may be a pale appearance with satellite projections into the surrounding liver tissue but some tumours have a homogeneous appearance and merge with the normal cut appearance of the parenchyma. Histologically the tumour cells resemble the normal polygonal hepatocytes and often contain inspissated bile between cells especially when well differentiated. The most commonly used classification of HCC is that of Gibson and Sobin employed by the World Health Organization.

- *Trabecular or sinusoidal* – tumour cells grow in cords or lamellae of variable thickness and are separated by sinusoids.
- *Pseudoglandular or acinar* – in addition to the basic trabecular pattern, areas of the tumour exhibit glandular pattern and canaliculi with or without bile plugs.
- *Compact* – tumour cells grow in solid masses compressing the sinusoids, which thus become indistinct.
- *Scirrhous* – cords of tumour cells are separated by fibrous septa.

All HCCs (and secondary tumours) lack a Kupffer cell population. This is important because some imaging tests, e.g. MRI using superparamagnetic ferric oxide contrast agents, are based on this characteristic. Various histological subtypes are recognized: the most common is the clear cell variant which accounts for up to 10%. The tumour cells appear 'clear' because they are laden with glycogen or lipid in the foamy cytoplasm and are sometimes associated with hypoglycaemia or hypercholesterolaemia respectively. The strong male prevalence seen with other HCCs is not shared by this histological subtype which has a male–female ratio of 1.6:1.0. Histological grades (I–IV) are used as an index of differentiation by some, but it has to be stressed that, in general, there is poor correlation between the histology and the prognosis of HCC. AFP-secreting tumours may show periodic acid–Schiff-positive staining.

HCC may spread to other segments of the liver. In this situation there is usually a large tumour mass with smaller separate satellite nodules. This is different from multifocal cancer where the nodules are of similar size. Spread through the liver is along the lumina of the hepatic and portal veins with the formation of tumour thrombus, which then embolizes to other parts of the liver. HCC is a highly vascular tumour as the neoplastic cells promote angiogenesis. Even tumour thrombus within the portal veins becomes vacularized by arterial tumour vessels and this mechanism is held responsible for the development of arteriovenous shunts that characterize this tumour. HCC can grow into and obstruct the biliary tract causing jaundice (icteric hepatoma).

Fibrolamellar HCC is a distinct variant with specific histological, histochemical and clinical features that occurs in young adults. This tumour has a better prognosis with cures rates of 50% after resection. The histology shows large polygonal cells in a dense fibrous stroma that forms bands or lamellar structures. The histology is similar to that of FNH and differentiation by imaging between the two may be difficult. The current consensus is that the two are not related aetiologically and that the FNH represents a hyperplastic response to the HCC. The AFP is not elevated in fibrolamellar carcinoma, but the serum vitamin B_{12} binding is usually elevated and this has been suggested as a tumour marker for fibrolamellar HCC. Also, some patients have elevated plasma neurotensin.

The pattern of recurrent disease after apparently successful resection suggests that direct infiltration along the hepatic veins and suprahepatic vena cava is common. Lymphatic spread to the portal tract is also common and distal lymphatic metastases are seen later in the natural history. Bloodborne metastases in the lungs are common.

Rarer malignant hepatic tumours

Hepatocellular cholangiocarcinoma combines histological features of both types of tumour and probably represents a coincidental occurrence. *Cystadenocarcinoma* occurs as a large lesion usually in adults. Microscopically, cystic spaces are lined in part with cuboidal epithelium with papillary projections.

Angiosarcoma (malignant haemangioendothelioma) consists of proliferating endothelial cells with fibrotic and haemorrhagic areas and large cavernous sinuses lined with dedifferentiated endothelium. The tumour cells may express factor VIII-related antigen. The tumour usually forms multiple nodules throughout the liver rather than a single mass. There is an established association with vinyl chloride production (as soluble monomer and solid PVC) or exposure to arsenic, Thorotrast or anabolic steroids. Angiosarcoma can occur in both children and adults. It usually metastasizes within the peritoneal cavity and carries a very poor prognosis. Presentation with life-threatening intraperitoneal haemorrhage is common. The related *epithelioid haemangioendothelioma* also expresses factor VIII-related antigen and, although it has a somewhat more favourable prognosis, it too can bleed intraperitoneally and develop extrahepatic metastases.

Adult hepatoblastoma is now referred to as malignant hepatic mixed tumour. Although composed of both epithelial and mesenchymal elements, these tumours are different from teratomas. They can form very large tumours.

Hepatic tumours in childhood

Primary liver tumours account for about 15% of abdominal tumours in childhood.

They include:

- hepatoblastoma
- HCC
- infantile haemangioendothelioma
- undifferentiated embryonic sarcoma (embryonal rhabdomyosarcoma)
- mesenchymal hamartoma
- malignant rhabdoid tumour
- mature teratomas.

Hepatoblastoma is the most common liver tumour in children and usually occurs below the age of 5 years with a male predominance. This is an aggressive, rapidly enlarging tumour. Metastases are present in 50% of patients at the time of presentation. The right lobe of the liver is involved in 70%. Various histological types are recognized: HB epithelial, HB mixed epithelial and mesenchymal, and HB NOS (not otherwise specified). The HB mixed type includes the teratoid hepatoblastoma which contains elements of the three germ layers. Irrespective of type, the AFP is elevated in the majority of cases. Large tumours may rupture with massive intraperitoneal haemorrhage that is often fatal.

Infantile HCC has similar morphological features to the adult type except for the absence of chronic liver disease/cirrhosis. It may occur in infants. Fibrolamellar HCC is encountered in older children

Undifferentiated embryonal sarcoma is usually diagnosed in children between the ages of 6 and 10 years and shows no male preponderance. Histologically the tumour consists of undifferentiated spindle cells and multinucleated giant cells in loose connective tissue. Differentiated components (cartilage, osteoid, striated muscle, etc.) are infrequent. Where possible, liver resection is indicated and about one-third of patients survive long term.

Mesenchymal hamartoma may form a lobulated multicystic or solid, at times pedunculated mass usually in the right lobe of the liver. It presents in infants (usually about 10 months), but may develop in older children, as a progressively enlarging mass. The tumour consists of epithelial line structures resembling bile ducts in a fibrous/myxoid stroma. The lesion is benign and cure is achieved by resection.

The rare *malignant rhabdoid tumour* is a very aggressive tumour consisting of sheets of cells containing eosinophilic intermediate cytoskeletal filaments. It is almost invariably fatal.

Clinical features of hepatic tumours

The incidence of HCC shows wide geographical variation with high-prevalence regions (Mozambique, South Africa, Far Eastern countries) and low-prevalence countries (1–2/100 000) such as the USA and UK. Some races seem to be at a higher risk, e.g. black population of South Africa and Chinese. HCC exhibits a male dominance with an average sex ratio of 3:1. In the early stages of the disease, HCC is asymptomatic; and is only discovered by screening (by ultrasound and AFP) in individuals at risk of the disease.

The predominant symptom results from an abdominal mass, which produces a dragging sensation on exercise. Other symptoms include anorexia, weight loss, abdominal or chest pain, vomiting, fever and, more rarely, changes in bowel habit and weakness. Jaundice and peripheral stigmata of chronic liver disease may be present in patients with HCC in a cirrhotic liver or arising in childhood biliary atresia. Also, jaundice may supervene from the growth of HCC into the bile ducts (icteric hepatoma). Some liver tumours at all ages present acutely with rupture and massive intraperitoneal bleeding.

Physical findings include obvious abdominal distension due to hepatomegaly or the presence of a hepatic mass. A bruit is heard in about 10% of patients. Ascites is common and is sometimes bloodstained. Additional infrequent clinical features include hypoglycaemia, hypercalcaemia, hyperlipidaemia and hyperthyroidism. Laboratory studies frequently show abnormalities of LFTs, but these may reflect underlying chronic liver disease. Haematological abnormalities may include anaemia due to intratumoral haemorrhage or polycythaemia due to anomalous erythropoietin release. Serum AFP levels are elevated in about one-third and are a useful cancer marker after resection. However, it is more commonly elevated in the US, Chinese and African populations. Serum AFP is more likely to be elevated in undifferentiated cancers and is a valuable marker for screening the cirrhotic population for HCC development. The reported sensitivity of serum AFP in the detection of HCC in patients with cirrhosis varies from 29% to 87%. Despite this, a rising level of AFP or a level greater than 500 ng/mL in patients with chronic active hepatitis or cirrhosis should be followed by real-time ultrasonographic examination of the liver and other imaging techniques to exclude HCC.

Hepatitis B and C should be looked for in all patients. For patients undergoing surgery a study of the coagulation parameters is necessary, as these may be abnormal and certainly will be affected by a major hepatic resection. Rarely there is hyperlipidaemia or hypercalcaemia with demineralization of the skeleton and spontaneous fractures.

Tumour localization and evaluation

Most patients are initially evaluated by ultrasound scanning to demonstrate size and position of lesions. More detailed information is obtained by helical high-dose CT or MRI, which reveals lesions greater than 1.0 cm in size. A chest radiograph may suggest direct diaphragmatic involvement or show pulmonary metastases. Both ultrasound and CT-guided biopsy are practised to obtain histological confirmation, but this practice is ill advised in patients with resectable tumours, as it encourages local implantation and may thus jeopardize cure. Occasionally all imaging modalities (ultrasound, spiral CT, MRI) fail to demonstrate a lesion in a cirrhotic patient who has other indications of the presence of an HCC, e.g. raised AFP. In these cases, laparoscopy with laparoscopic contact ultrasonography may be helpful. For large posteriorly placed tumours, cavography is essential to exclude invasion of the inferior vena cava but simple compression or deviation of the inferior vena cava does not preclude a successful resection. There is probably no place for a trial dissection of a HCC and the surgeon should have fully assessed the probability of a successful resection before embarking upon it.

With the advent of monoclonal antibodies, newer immunoscintigraphic methods are being developed. These include labelled antibodies against AFP and antibodies to surface antigens of human HCC (XF-8, AF-20). Preliminary results look promising.

Surgical treatment

Preoperative preparation

Disturbance of LFTs, particularly with regard to serum levels of bilirubin and albumin, may indicate the seriousness of liver damage in patients with chronic liver disease and tumour. Child's classification is used to judge the operative risk prior to resection. Patients classified as grade B or C have effectively lost 50–60% of their liver parenchyma. ICG clearance is used in many centres to evaluate the hepatic reserve and thus the ability to withstand major resection. The arterial ketone body ratio (KBR) of acetoacetate/ -hydroxybutarate is a good indicator of the hepatic mitochondrial redox potential and for this reason is used in some centres in the preoperative evaluation. Resection is poorly tolerated in patients with a KBR <0.4.

In some patients with large tumours of the right liver requiring an extended right hepatectomy (segments IV–VIII), the residual left lobe (segments II and III) may be too small and in these patients resection will be followed by acute liver failure even if the liver parenchyma is healthy. In these patients preoperative right PVE (see below) is used to achieve substantial hypertrophy of the left lobe before the resection is undertaken 4 weeks later. Patients with normal liver parenchyma will regenerate the liver bulk but patients with cirrhosis are not able to regenerate further liver tissue. Hence major resections are contraindicated in cirrhotic B and C patients.

Fluid and electrolyte disturbances should be corrected prior to surgery and vitamin K is given routinely. Nutritional supplementation with the maximum protein load tolerable by the patient may elevate the serum albumin though albumin infusions are usually required postoperatively. Patients should receive an intravenous glucose load (1 L of 10% dextrose) prior to surgery in order to ensure that hepatic glycogen is maximal.

Resection

Surgical resection remains the first-line treatment of HCC even in patients with cirrhosis, provided their hepatic reserve is good. The best results are obtained in patients with lesions less than 2.0 cm and in those with encapsulated well-differentiated tumours. The 5 year survival rate after resection averages 35–38%. Recurrence after resection is common and reaches 55% at 5 years. The risk of intrahepatic recurrent disease after curative resection is enhanced by the presence of a macroscopic portal thrombus at the time of surgery, indicating possible spread of cancer cells from the thrombus during the surgical intervention. Thus only 21% of intrahepatic recurrences are situated near the hepatic resection line and the majority (79%) are located away from the hepatic stump. For this reason, some surgeons advocate intraoperative embolization (with starch microspheres) of the portal branch supplying the tumour at the time of the resection.

The predictors of survival by univariate analysis are vascular invasion, advanced age, multiple tumours and lack of capsule but only vascular invasion remains significantly predictive on multivariate analysis.

The objective of surgical resection is to excise the lesion safely with a margin of healthy liver tissue (Figure 24.30). This can be achieved by careful operative assessment. Even so, intraoperative ultrasound is used to locate the lesion in relation to major structures (hepatic veins) and to ensure that no other satellite foci are present. Small liver tumours are amenable to segmental resection providing a tumour-free margin of at least 1.5 cm (Figure 24.31). This approach, i.e. segmentectomy, is most suited to patients with chronic liver disease where tumours have been detected relatively early by screening.

Large tumours require a formal dissection along major planes. Where the tumour is localized to segments V–VIII, a right hepatectomy is required. If it encroaches on segment IV, an extended right hepatectomy (including segment IV of the left liver) is necessary. Lesions in segments II and III are excised by left lobectomy but where, as is often the case, the tumour approaches the falciform ligament and segment IV, then a left hepatectomy (segments II–IV) is necessary. Some cancers extend across the

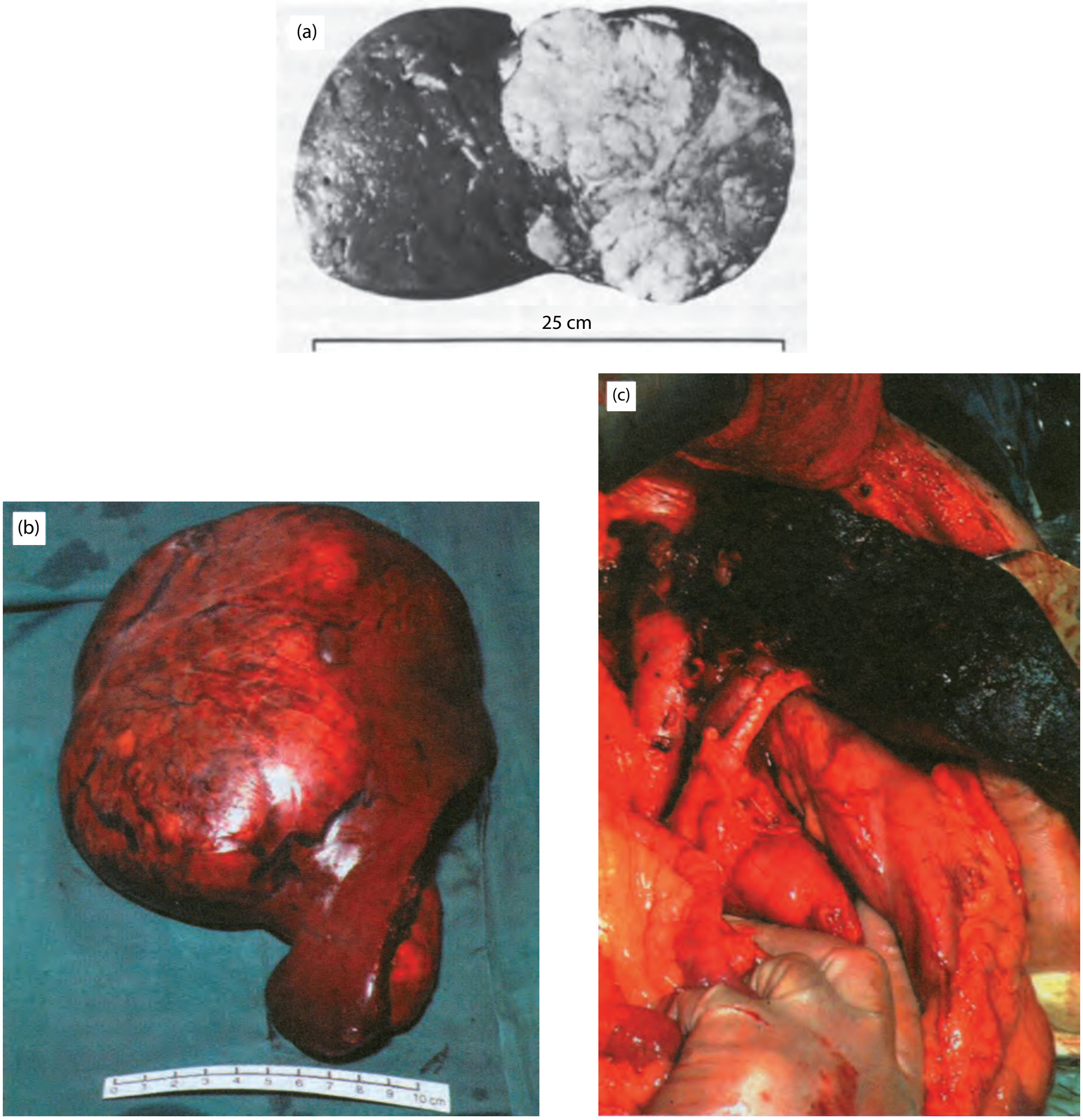


Figure 24.30 (a) Transected surgical specimen showing a large bulky tumour which appears lobulated and to have necrotic areas. There is evidence of satellite nodules in the periphery of the main tumour. (b) Surgical specimen after extended right hepatectomy (right trisectionectomy) and caudate lobectomy for a large hepatoma. (c) Intraoperative view at the end of the resection. The transected liver surface on the residual segments II-III has been treated with an argon-beam coagulator.

plane between the two livers particularly from the right into segment IV. In some patients requiring major resection for both primary and secondary hepatic tumours, e.g. especially extended right hepatectomy, the residual liver parenchyma (segments I–III) is too small. Avoidance of major blood loss during major hepatic resections underlies a good postoperative outcome. Bleeding is considerably reduced during the hepatectomy if the central venous pressure is reduced to around zero by ante-Trendelenburg position, epidural anaesthesia and vasodilator therapy.

Hepatic transplantation

The results of hepatic transplantation for HCC have improved with recurrence rates of 20% and survival rates of 45% at 5 years. The best results have been in cirrhotic patients with small tumours where partial resection could have been considered as an alternative. Thus, aside from occult small lesions (<2.0 cm) in patients requiring transplantation for chronic liver disease and cirrhosis, the indication for this surgical treatment is limited to a selected group of patients with fibrolamellar carcinoma. Even in this group, some consider that, if the tumour is too large to resect, the outcome following transplantation is not good.

Postoperative care after liver resection

Patients after liver resection require major parenteral support with diminishing daily requirements as liver regeneration takes place. Following hepatectomy/lobectomy there is transient portal hypertension and a sizeable sequestration of blood in the portal venous system.

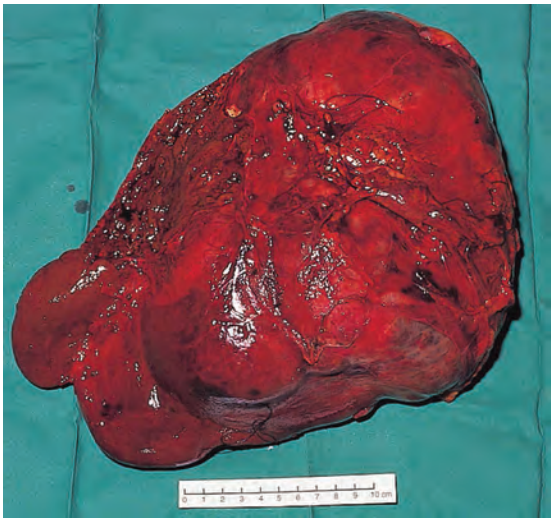


Figure 24.31 Bisegmentectomy V and VI for metastatic carcinoid tumour.

Jaundice is minimal and lasts about 1 week. More profound and persistent hyperbilirubinaemia may indicate bile duct obstruction and the need for further surgery. Hypoglycaemia can occur in the early postoperative course. Regular monitoring of blood glucose levels is thus necessary. Usually this complication can be avoided by infusing 5–10% glucose during and after operation. Intra- or postoperative coagulation defects can also be countered by the prophylactic use of fresh frozen plasma (at least two units daily during the first 4 days) and vitamin K injections. Since the half-life of albumin is 8–24 hours, hypoalbuminaemia is universal after a major hepatic resection and only partly correctable by plasma infusion. Repeated plasma or albumin infusion may be necessary for at least 1 week postoperatively until hepatic regeneration is sufficient to maintain plasma levels. Regeneration of liver documented by imaging occurs by 3 months if the parenchyma is normal.

Non-surgical management of primary hepatic cancer

Chemotherapy

Chemotherapy is used in some patients with unresectable lesions although the results are generally poor. The chemotherapy may be administered systemically or regionally with hepatic arterial infusion using either external or implantable pumps. Only one report has demonstrated that hepatic arterial infusion with floxuridine, doxorubicin and mitomycin C is associated with increased survival compared with systemic intravenous therapy, with the vast majority showing no difference. In general, a higher objective response is obtained by regional chemotherapy but this is offset by a higher rate of post-treatment complications (chemical hepatitis, biliary sclerosis, peptic ulceration and gastritis/duodenitis). A recent retrospective large series of chemotherapy for hepatic malignancy documented an overall response rate of 13% in patients with HCC. Thus it is doubtful whether this poor response justifies the morbidity and costs of this treatment. The results of adjuvant immunotherapy with cytokines such as interleukin (IL)-2 and interferons have been equally disappointing to date. Some response to -interferon can occur in patients with HCC arising on a background of cirrhosis due to hepatitis B or C.

Chemoembolization

TACE with gelatin sponge and/or ethiodized oil (with or without added chemotherapeutic agents such as cisplatin or doxorubicin) imparts some objective response and symptomatic improvement in patients with unresectable disease. TACE has also been used in the preoperative setting before surgical resection but there is no evidence that this combined approach improves the survival or reduces the recurrence rates.

Radiotherapy

Supervoltage radiotherapy has been used in unresectable disease but the treatment is limited by the dose-related radiation-induced hepatitis. Thus dosages greater than 30 Gy are not tolerated. Nonetheless, symptomatic relief of pain can be obtained by this therapy. A more recent approach is interstitial radiation using ^{131}I-labelled antibodies to ferritin and AFP with the aim of targeting the radiation dose to the tumour. The early results with this targeted interstitial radiotherapy seem promising, with 50% partial response rates.

Metastatic liver tumours

Direct invasion of the liver may result from locally advanced cancers of the stomach, pancreas and hepatic flexure of the colon. More commonly, hepatic metastases are the result of vascular spread from the primary tumour via either the portal vein or hepatic artery. The liver is by far the commonest site of metastatic disease from gastrointestinal, bronchial and breast cancers. In part this is due to its dual blood supply. Although many of the secondary deposits appear necrotic on cut section, there is sufficient blood supply to allow multiple deposits to expand, and livers weighing in excess of 10 kg at autopsy have been recorded. Although the histological picture normally reflects the primary tumour, there is a general tendency towards dedifferentiation.

Hepatic metastases are often asymptomatic, but patients with widespread involvement or large superficial deposits may suffer abdominal and back pain secondary to stretching of the Glissonian capsule or haemorrhage and necrosis within the tumour. Although many patients appear physically well when liver metastases are first detected, as the disease progresses, malnutrition, jaundice, ascites and cachexia are inevitable. By the time liver metastases become symptomatic, there is usually massive involvement of the liver. Thus the objective of modern management is the detection of early asymptomatic disease in patients at risk of secondary hepatic deposits, especially with colorectal cancer, when the results of surgery or *in situ* ablation are better and can even be curative.

Colorectal cancer liver metastases

Colorectal cancer is the commonest gastrointestinal malignancy and the second commonest cause of cancer death in the Western world. It is estimated to cause 57 100 deaths per annum in the USA and 17 000 deaths per annum in the UK. The liver is usually the first site of metastatic disease and may be the only site in 30–40% of patients with advanced cancer. At the time of initial diagnosis of colorectal cancer, 20–25% of patients will have clinically detectable liver metastasis. A further 40–50% will develop liver metastases, most commonly within the first 3 years of follow-up after resection of the primary tumour.

The median survival of untreated colorectal liver metastasis (CRLM) is around 6–8 months, varying with the extent of disease at presentation. Although overall survival of patients with CRLM has improved with modern chemotherapy agents and median survival of 18–20 months can be achieved with the most aggressive protocols including 5-fluorouracil (5FU), oxaliplatin and irinotecan, surgery remains the only treatment modality that offers the prospect of cure. A recent systematic review has shown that 30–40% of patients will achieve 5 year survival after liver resection and, although relapse may still occur between 5 and 10 years, 20% will still be alive at that point. Further relapse is unlikely after 10 years. Some tertiary referral centres have achieved 5 year survival in excess of 50% after liver resection in subgroups of patients with more favourable prognostic factors, and survival has continued to improve steadily during the last two decades. A recent epidemiological study in the UK showed that the outlook for patients with CRLMs who had surgical resection was no different from that of patients with primary colorectal tumour in stage III (Dukes's C). Until recently, however, only 10–20% of patients were considered suitable for attempted curative resection with the remaining patients being offered palliative and symptomatic treatments.

Criteria for resection

In the past, the decision to resect CRLMs was relatively straightforward. Liver resection was considered appropriate only in patients who had one to three unilobar metastases, preferably presenting at least 12 months after resection of the primary tumour, whose disease was resectable with at least a 1 cm margin of healthy liver tissue and who had no hilar lymphadenopathy or extrahepatic disease.

More recently, with advances in surgical technique, preoperative and postoperative care and chemotherapy, indications for liver resection have expanded and even patients traditionally expected to have poor prognosis have been reported to have long-term survival following liver resection. A shift has occurred in the criteria used for assessing resectability, from mere morphological criteria to new ones based on whether complete resection (R0) can be achieved. Instead of resectability being defined by what is removed, resectability should now be determined by what will remain. CRLMs are nowadays considered resectable if (1) the disease can be completely resected, (2) two adjacent liver segments can be spared with adequate vascular inflow and outflow and biliary drainage and (3) the future liver remnant (FLR), calculated on preoperative CT scan with volumetry software, is adequate. The safe FLR will vary from patient to patient depending on body mass index and underlying disease. Patients with an otherwise normal liver can survive with a FLR as low as 20–25% of the initial volume; a FLR of 30% is desirable following chemotherapy and this increases to 40% in the presence of chronic liver disease. Some conditions remain relative contraindications to surgery like invasion of a first-order division branch of the hepatic pedicle, contact with the contralateral branch, contact with the inferior vena cava, invasion of all three hepatic veins, presence of coeliac trunk lymph nodes and presence of non-treatable extrahepatic disease.

The presence of bilateral metastases has also traditionally been considered as a contraindication to surgical treatment. More recently, however, resections in two or more stages have been successfully performed for bilobar disease with good short- and long-term results. In a 'staged resection', one lobe of liver is resected or otherwise cleared of disease first followed by a period of regeneration, usually 2–3 months, after which the contralateral side is operated on. This strategy can be adopted for bilateral lesions in which a one-stage resection would leave an insufficient liver remnant with consequent liver failure or where the metastatic disease cannot be completely cleared with a single resection. Radiofrequency ablation or other techniques of *in situ* liver ablation have also been used in combination with resection surgery with the aim of increasing resectability and temporarily controlling disease progression. In this clinical setting, PVE and two-staged resection can often be used together with PVE employed to induce hypertrophy of the future remnant following clearance of metastatic disease (see section Embolization of portal vein prior to major hepatectomy).

When assessing resectability, it is important to observe that a positive liver resection margin remains a predictor of poor prognosis. However, although the minimum resection margin has historically been >1.0 cm, this has significantly reduced in recent years, and recent studies have shown that the width of surgical margin has no effect on long-term survival and even a margin of 1 mm is sufficient and is associated with long-term survival. A microscopic resection margin of 1 mm is nowadays considered acceptable, particularly if the liver transection has been performed with modern surgical devices like CUSA (Cavitron Ultrasonic Surgical Aspirator) or radiofrequency devices. It should also be recognized that several millimetres worth of resection margin may be removed with the CUSA and that a radiofrequency ablation of the transected liver surface for a depth of several millimetres (0.2–0.8 mm) is achieved with devices like Tissuelink or Aquamantis (Salient Surgical), increasing the safety of a smaller resection margin.

Synchronous metastases

At the time of diagnosis with colorectal cancer, 15–25% of patients already have liver metastatic disease (synchronous metastases). These patients pose a significant therapeutic challenge and the optimal treatment strategy is still uncertain. Recent data, however, indicate that a simultaneous colon–liver resection is technically feasible, safe, cost-effective and entails additional patient benefits in terms of psychophysical recovery and quality of life. It has been demonstrated that a resection of fewer than three hepatic segments can be performed in association with any type of colorectal resection without any increase in mortality or morbidity and with significant reduction in total operative time and total hospitalization. The combination of a major hepatectomy (four or more segments) with a colorectal resection seems however to be associated with an increased mortality and therefore remains unadvisable.

Survival at 3 and 5 years after a synchronous colon–liver resection has been found to be improved in a recent meta-analysis when compared with a classic procedure in two stages; an increased number of liver recurrences has however been observed at 1 year, caused by occult liver disease becoming overt and by the lack of patient selection operated by the time interval between colectomy and hepatectomy, where up to 30% of patients can show progressive disease even despite chemotherapy and hence never proceed to liver resection.

Embolization of portal vein prior to major hepatectomy (trisectionectomy)

Removal of more than 60–65% of the liver parenchyma in adults carries an increased risk of liver failure and mortality, particularly in obese patients, postchemotherapy patients and patients with underlying chronic liver disease. This problem is quite common in patients undergoing an extended right hepatectomy (trisectionectomy) and in some patients undergoing a right hepatectomy where only a very small left lateral sector (left lobe – segments II and III) is present.

Preoperative PVE, by redistributing the portal venous blood in the liver, induces atrophy of the liver to be resected and hypertrophy of the prospective liver remnant. PVE is usually well tolerated with virtually no symptoms and barely measurable liver function disturbance. Occasionally, however, some may develop liver failure or ascites. This technique increases the FLR thus significantly expanding the pool of resectable patients: up to 85% of patients previously judged irresectable on the basis of an insufficient FLR can undergo attempted major hepatectomy after PVE while only 15% had to be definitively turned down (inadequate hypertrophy of FLR or tumour progression being the most common reasons), as shown in a recent meta-analysis.

Notes on imaging and staging

CT

Contrast CT of the chest, abdomen and pelvis is commonly used as the initial imaging modality. Iodinated contrast allows characterization of hepatic lesions based on enhancement patterns. During the portal venous phase, normal liver enhances uniformly while liver metastasis appear hypodense. This can be attributed to the largely arterial blood supply of these lesions. Smaller lesions may appear more evident in images acquired during the arterial phase scan; however, for subcentimetre lesions, MRI has a higher sensitivity.

MRI

MRI has several advantages over CT such as no risks from radiation exposure and no adverse reactions to iodinated contrast agents. CRLMs are hypo- to isointense on T_1 and iso- to hyperintense on T_2 weighted images. MR contrast agents provide critical tumour characterization and can be safely used in patients with iodine contrast allergy and renal failure. Other agents, including gadolinium chelates or iron oxide agents, may provide additional benefits in selected applications. The degree and nature of tumour vascularity form the basis for liver lesion characterization based on enhancement properties. It is particularly useful in evaluating indeterminate subcentimetre hepatic lesions and lesions in livers with fatty infiltration.

CT-PET

This is commonly used in patients with liver metastasis being considered for resection. PET is an imaging modality which images metabolic activity via the distribution of positron-emitting tracers that are incorporated into metabolic processes. In practice, this translates into the parenteral administration of a metabolically radioactive (positron emitting) tracer into the body which is then incorporated into a metabolic pathway. The emissions of positrons are then imaged with the PET scanner. CT-PET combines the advantages of CT with the functional ability of PET, by the fusion of PET images and CT images acquired at the same time, helping to accurately localize the area of increased metabolic activity. This modality is highly sensitive however and any focal area of hypermetabolism can give a false-positive result. It is also expensive and has a questionable role in determining lesions less than 1 cm in size.

Laparoscopy

Laparoscopy with or without intraoperative ultrasound has also been used in the preoperative assessment. Its role has evolved as radiological detection has improved. It has high sensitivity in detecting surface CRLM and peritoneal carcinomatosis. Overall, the yield of laparoscopy for metastatic colorectal cancer ranges from 6% to 36%, with 10% in the largest published series. The yield of detection has been shown to be related to duration of disease-free survival, number of hepatic lesions, size of hepatic lesions, nodal stage of primary tumour and prehepatectomy carcinoembryonic antigen level (<200 ng/mL) as part of a clinical risk score assessment.

Operative techniques

Transection techniques

Blood loss during liver resection is associated with increased postoperative complications. Technological innovations in the transection of liver parenchyma have been developed with the aim of reducing blood loss to which reduced morbidity, mortality and safety can at least partially be attributed. Haemorrhage during hepatotomy can be controlled or prevented by a Pringle manoeuvre, which temporarily interrupts all blood inflow to the liver, although this may be associated with ischaemia–reperfusion injury. Reduction of the central venous pressure, with continuous infusion of vasodilating drugs, like glyceryl trinitrate or epidural anaesthesia, has also been advocated to reduce blood loss during hepatic resection because the hepatic veins are in direct continuity with the inferior vena cava.

Numerous techniques and devices are now available for the transection of the liver parenchyma. Historically, liver resections were performed with the finger fracture technique, whereby the liver was crushed between the operating surgeon's fingers preserving the robust fibrous structures which would then be divided in turn. This brutal and imprecise technique was then replaced by the crush-clamping techniques and more recently by the newer and more technologically advanced techniques involving often expensive surgical devices like radiofrequency dissecting sealers, CUSA, hydrojet and others.

With meticulous transection technique and with the help of these devices, hepatic resections are now routinely performed with minimal blood loss in specialized centres and without the need of routine Pringle clamping.

For additional notes on the surgical strategy during liver resection see section Techniques of anatomical liver resection and intraoperative ultrasound.

Chemotherapy

Historically chemotherapy has been reserved for patients with unresectable CRLM where its survival benefits are well documented. Response rates of 48–78% have been reported with protocols combining multiple agents like 5FU, oxaliplatin, irinotecan, with median survival ranging from 16 to 31 months and survival at 3 years of 44–55% in selected groups.

More recently however these chemotherapy regimens have also been applied in neoadjuvant settings prior to liver resection in cases where resectability was borderline or downstaging desirable. It has been demonstrated that some 10–30% of patients with unresectable disease can be downstaged successfully and become suitable for definitive surgery with 5 year survival of 30–35%. Indeed, conversion to operability is now being considered a surrogate endpoint in trials of stage IV colorectal cancer.

The role of perioperative chemotherapy in outright resectable liver metastases remains controversial despite the results of a recent RCT (EPOC) and no definitive benefits have been demonstrated yet. Patients who received FOLFOX4 had improved 3 year progression-free survival (36.2% vs 28.1%) compared with patients who had surgery alone, but this came at the price of a reduction in resection rate in the chemotherapy group, mainly caused by tumour progression, poor patient condition and patient refusal (overall 12.1% vs 4.4%). Furthermore an increase in postoperative complications (25.2% vs 15.9%) was observed after neoadjuvant chemotherapy. Overall, despite a 7–9% increase in survival in resected patients who received neoadjuvant chemotherapy, these results do not seem to justify the routine adoption of it in outright resectable liver metastases and more evidence is needed.

It would appear sensible, however, despite the current lack of absolute data, to routinely offer postoperative chemotherapy after liver resection, which may yield the same survival benefits of neoadjuvant treatment without determining increased surgical morbidity and risking disease progression. After all, the survival benefits of adjuvant chemotherapy for colorectal cancer with lymph node metastases are well established and it is reasonable to assume that similar benefits may be achievable in patients with liver metastases.

The spreading of neoadjuvant chemotherapy presents challenges as well as benefits. Chemotherapy-associated steatohepatitis has become increasingly common and its consequences are well recognized. This refers to the bluish discoloration, oedema and spongiform consistency of the liver similar to that in early cirrhosis. In studies which have compared the short-term outcomes of patients who received chemotherapy prior to resection to those patients who did not, significantly more postoperative complications have been observed in the chemotherapy group. Moreover, the postoperative morbidity is associated with the number of chemotherapy cycles.

Monoclonal antibodies

Cancer cells can proliferate through the uncontrolled production of growth factors, or an increased number of growth factor receptors on their cell membranes. For example, epidermal growth factor receptor (EGFR) is overexpressed in around 75% of colorectal cancers, with K-ras one of the better known downstream pathways. Mutation and upregulation of this pathway is associated with increased cell proliferation. Conversely, inhibition reduces the growth of cancers expressing this receptor. Such activation or inhibition of EGFR is dependent on normal or wild-type K-ras expression. However, mutations in the *KRAS* gene, found in 30–50% of patients with colorectal cancer, can lead to cell proliferation that is independent of the action of EGFR, hence they show no response to EGFR inhibitors.

The growth of primary tumours and metastases also depends on the formation of new blood vessels (angiogenesis) to supply these tissues once they reach a critical size. Vascular endothelial growth factor upregulation has been found in many human tumours, including colorectal cancer, and is important in tumour angiogenesis. However, it also plays a crucial role in liver regeneration in patients undergoing liver resection.

Cetuximab

A recent meta-analysis analysed data from seven RCTs of chemotherapy with or without cetuximab, involving a total of 4617 people with metastatic or advanced colorectal cancer. It found that cetuximab increased the likelihood of a response to treatment and also improved progression-free survival. One study of 1217 previously untreated patients received chemotherapy with or without cetuximab. The median progression-free and overall survival times were said to be longer with cetuximab (8.9 vs 8.0 months; and 19.9 vs 18.6 months, respectively). The rate of resection of liver metastases after chemotherapy has been used as a surrogate outcome measure. Cetuximab increased the number of patients who had resection of their liver metastases (7.0% vs 3.7) and the rate of R0 resection with curative intent before disease progression (4.8% vs 1.7%). The National Institute for Health and Clinical Excellence (NICE) recommends that for patients with the wild-type *KRAS* gene cetuximab is used with chemotherapy as first-line treatment for metastatic colorectal cancer. In addition, the following criteria must also be met: the primary colorectal tumour has been resected or is potentially operable; the metastatic disease is confined to the liver and is unresectable; and the patient is fit enough to undergo surgery to resect the primary colorectal tumour and to undergo liver surgery if the metastases become resectable after treatment with cetuximab.

Bevacizumab

A meta-analysis of data from five RCTs, involving 3103 patients with metastatic colorectal cancer, assessed chemotherapy with or without bevacizumab. Compared with the combination treatment, chemotherapy alone was associated with shorter progression-free survival (median 7.1 vs 9.7 months) and overall survival (median 17.7 vs 20.5 months). Only one of the studies included in the meta-analysis provided information on the number of participants whose liver metastases became resectable (8.4% with bevacizumab vs 6.1% without). NICE guidance, published in December 2010, has not recommended bevacizumab in combination with oxaliplatin-containing regimens, as it was not considered to be a cost-effective use of NHS resources for first- or second-line or later treatment of patients with metastatic colorectal cancer.

In situ ablative therapies

In essence, *in situ* ablation consists of local destruction of the tumour without resection. This is achieved by thermal methods aimed at heating the tumour to a temperature above 42.5°C for several minutes or by rapid freezing to −40°C or by chemical denaturation. The various ablating technologies include:

- rapid freezing – cryotherapy
- radiofrequency heating probes
- microwave heating
- interstitial laser hyperthermia
- heating by high-intensity focused ultrasound
- chemical ablation by alcohol injection; biochemical ablation by p53-induced apoptosis (developmental).

Currently *in situ* ablation of liver tumours is used

- when the disease is inoperable either because of multiple deposits in both lobes, or in the case of HCC because the functional reserve of the liver precludes safe resection (Child B, C)
- in conjunction with resection for bilateral disease
- whenever the patient comorbidities would not allow a safe hepatic resection.

In situ ablation can be carried out at open surgery, percutaneously or via the laparoscopic approach. The major advantage of the percutaneous and laparoscopic routes is that the procedure can more easily be repeated several times. This is an important consideration in patients with metastatic liver disease, as the majority of these patients develop fresh lesions in time. The laparoscopic approach carries advantages over the percutaneous technique – more accurate ablation and reduced risk of collateral damage to adjacent organs. The actual ablation is monitored by laparoscopic contact ultrasonography.

The initial results of phase II and III studies comparing *in situ* ablation with surgery are promising but the experience is still limited. It appears however that, for lesions of a maximum 3–3.5 cm in diameter, *in situ* ablation may well achieve long-term survival similar to surgery in case-matched settings and in experienced hands. For larger lesions, the risk of incomplete ablation and recurrence seems to be higher and therefore long-term results are worse.

Radiofrequency ablation

With this modality, a high-frequency alternating current (100 000–500 000 Hz) is passed through the tissue to cause intense ionic agitation resulting in frictional heating. Normal and neoplastic tissue is destroyed if heated to above 43°C for 5 minutes. Multielectrode probe systems increase the ablative power of radiofrequency generators. Even so the ablative zone rarely exceeds 3.5 cm. This is because of charring around the electrodes, which increases the impedance and thus limits current flow in the tissue. This problem has been overcome by the introduction of 'cool-tip' electrodes. These are hollow and are kept cool (below 30°C) by closed cold water circulation. Because of the absence of charring around the electrode during use, the ablative capacity is considerably enhanced (Figure 24.32). Radiofrequency thermal ablation has been used in patients with inoperable primary and secondary hepatic tumours. Most of the reported experience has been on the use of radiofrequency ablation by the percutaneous route under radiological or external ultrasound guidance. The laparoscopic/contact ultrasound approach carries a number of potential distinct advantages over the percutaneous route. In the first instance the positive pressure pneumoperitoneum by reducing liver blood flow decreases the heat sink, resulting in a larger ablative zone. Second, the approach permits precise visual and contact ultrasound-guided insertion of the electrode in the hepatic lesion; and, third, the risk of collateral damage is minimized.

Microwave

This is similar in principle to the above but uses microwave energy instead. It seems to be accompanied by a higher morbidity rate, especially of biliary fistula, than radiofrequency ablation. It is popular in Japan, where it is also used during open surgery for inoperable liver tumours.

Cryotherapy

Cryotherapy is the oldest *in situ* ablation modality and has been used in the treatment of inoperable liver tumours for more than 25 years. The principle is very rapid freezing followed by a slow thaw. The tumour must be frozen to −40°C for complete destruction. Modern cryo-units use high-efficiency liquid nitrogen (boiling temperature −190°C) recirculating through implantable probes (2–4 mm) that are impaled in the lesion under ultrasound control. The objective is to achieve an ice-ball that totally encompasses the tumour and extends into the surrounding normal parenchyma for at least 1.0 cm (Figure 24.33).

Non-colorectal cancer liver metastases

The recent success of an aggressive surgical approach to the management of CRLM has, in part, provided the impetus for stimulating the use of liver resection for non-CRLM disease. Extrapolating surgical strategies from one malignancy to another is reasonable in some cases; however, fundamental biological differences between various neoplasms dictate careful consideration and understanding of the tumour's

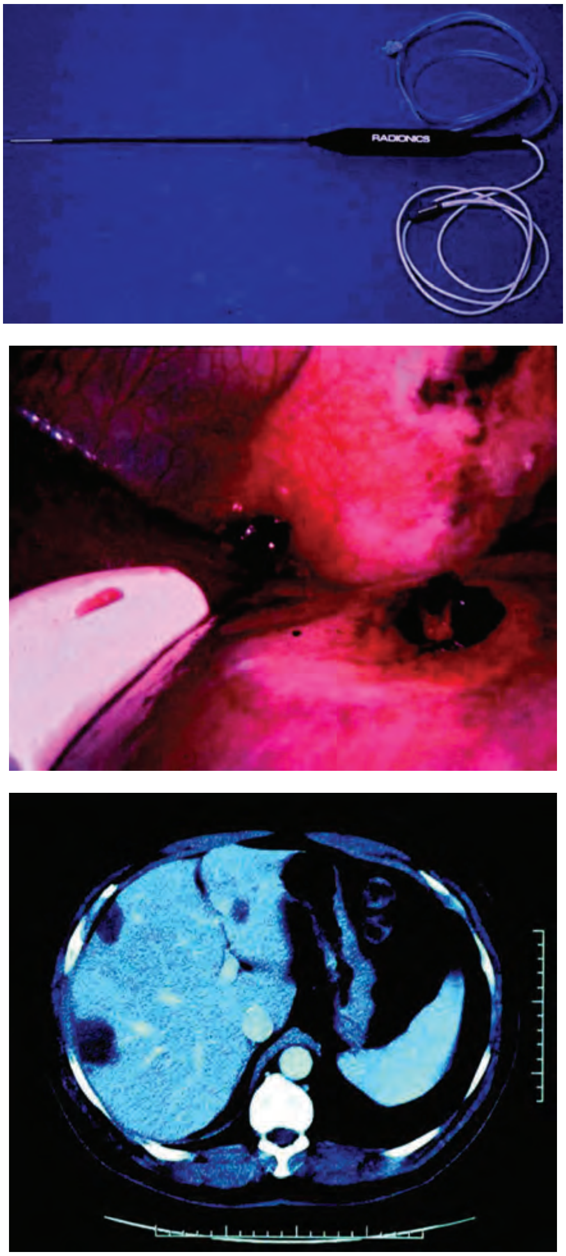


Figure 24.32 (a) Cool-tip electrode for radiofrequency ablation (Radionics); (b) laparoscopic thermal ablation of secondary hepatic deposits; (c) postoperative CT scan showing satisfactory incorporation of ablated tumours (dark areas in liver parenchyma).

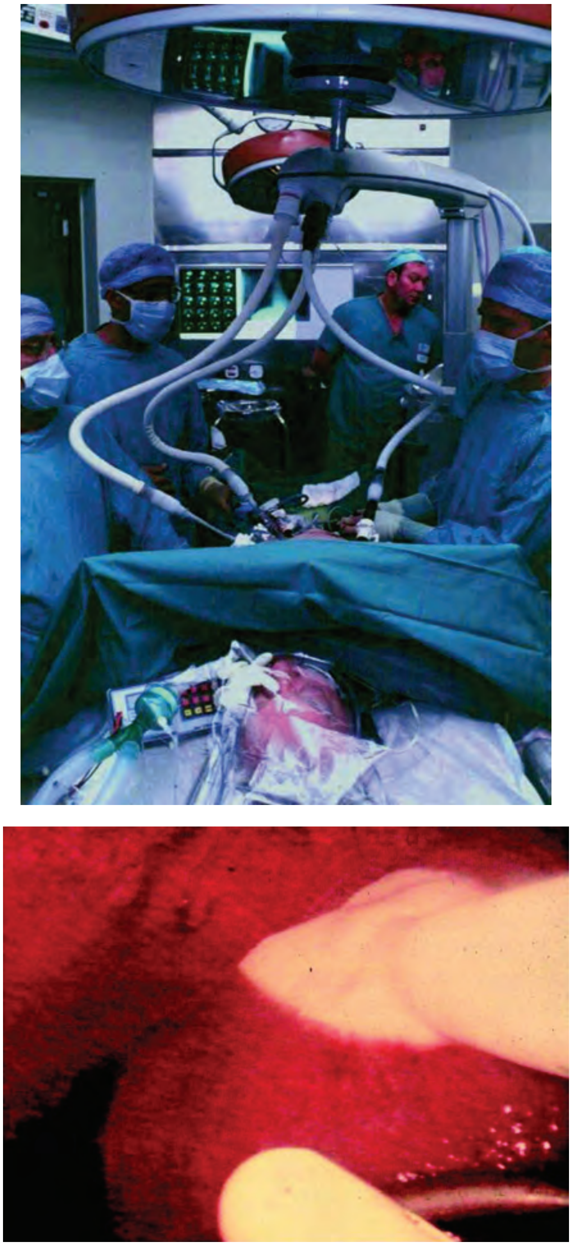


Figure 24.33 (a) Laparoscopic cryoablation of metastatic liver disease. (b) The ablation is monitored by laparoscopic contact ultrasonography – the expanding ice-ball has a hyperechoic margin that surrounds an acoustic shadow.

natural history and prognosis. Treatment decisions must take into account clinical surrogates of tumour biology. Patients with synchronous liver metastases or short disease-free interval or with extrahepatic disease may have more aggressive tumours and may be less likely to gain significant survival benefit from liver resection. In most cases, liver resection should be performed with curative intent and the ability to achieve negative resection margins is a significant prognostic factor. The majority of patients with non-CRLMs have disseminated disease and are not candidates for hepatectomy. Exceptions include liver metastases for NETs, epithelial ovarian cancer and testicular malignancies, where debulking is considered useful as a palliative manoeuvre to improve overall survival. Ablative strategies and systemic or locally delivered chemotherapy can be used as adjuncts to resection. Radiofrequency ablation has been reported to be safe and successful at achieving local control in patients with breast cancer, ovarian cancer and NETs. Its major limitation is the inherent difficulty of achieving complete necrosis in tumours larger than 3 cm.

Transarterial embolization (TAE) and TACE take advantage of the differential blood supply of liver metastases, which depend mainly on the hepatic arteries, and the normal parenchyma, which relies more heavily on the portal circulation.

Breast

The case for resection of breast cancer metastases is evolving, with some liver surgeons advocating resection as one of various cytoreductive therapies for this disease. Response to chemotherapy appears to be an important predictor of overall survival following liver resection for metastatic breast cancer. In the largest series, patients whose disease remained stable or progressed during prehepatectomy chemotherapy were 3.5 times more likely to die than responders. Preoperative chemotherapy has been recommended in all cases, with hepatectomy reserved for those who show objective response.

Ovarian

Although the liver is rarely the only site of metastatic disease in ovarian cancer, hepatectomy can be an important component of a primary cytoreductive strategy. Ovarian cancer can involve the liver through the development of peritoneal lesions on the surface of the liver (stage III) or intraparenchymal (stage IV). Survival is improved for patients with stage IV disease who have undergone adequate debulking surgery including hepatectomy.

Gastrointestinal stromal tumours

Gastrointestinal stromal tumours (GISTs) represent a group of mesenchymal malignancies that originate from the interstitial cells of Cajal of the gastrointestinal tract. Primary GISTs arise in the stomach (60%), small intestine (25%), colon and rectum (10%), with the remainder found in various other sites (gallbladder, appendix, omentum, mesentery). Surgery is the treatment of choice for resectable primary GISTs. Commonly, the primary tumour is classified into four prognostic categories, ranging from very low to high risk, according to size of lesion and the number of mitotic figures identified. Between 40% and 60% of resected patients with primaries in the intermediate- to high-risk categories recur with two patterns of recurrence: local recurrence with peritoneal disease or intraparenchymal liver metastases. Liver metastases are frequently multiple, large and bilobar making curative resection unlikely. The treatment of unresectable primary and metastatic GISTs has been revolutionized by imatinib, a selective tyrosine kinase inhibitor. Most patients with metastatic GISTs will receive imatinib as first-line treatment with a clinical response in 80%. The response is durable, with a reported median survival greater than 24 months. Eventually, though, many patients develop imatinib resistance and disease progression. Second-line agents (sunitinib and nilotinib) are currently being tested.

Neuroendocrine tumours

Gastrointestinal NETs are a diverse group of tumours that originate throughout the gastrointestinal tract and can be classified into carcinoid and pancreatic histological subtypes. Most NETs of gastrointestinal origin demonstrate indolent growth. Nonetheless, the majority will have disseminated disease at the time of diagnosis and 5 year survival is 50–80%. NETs metastasize preferentially to the liver, and in many patients the liver remains the only site of metastatic disease for a prolonged period of time. The majority of patients have multifocal, bilobar disease and more than half will have >50% liver involvement. Liver resection may be performed with curative intent, for symptom control or prolongation of survival in the palliative setting. A review article demonstrated that hepatic resection for metastatic NETs resulted in almost double 5 year survival rates over the unresected group (30–40% to 47–82% respectively). In addition, cytoreduction offers the most effective and durable palliation of symptoms. Radiofrequency ablation used in isolation can achieve symptomatic relief and local control in 60–80% of NETs with liver metastasis, but the duration of response is variable. TAE and TACE have also been shown to achieve reasonable palliation for very bulky or symptomatic unresectable tumours. However duration of response can be short and repeat embolizations are often required.

Laparoscopic liver surgery

Since the introduction of laparoscopic cholecystectomy in 1987, the laparoscopic approach has been applied to the full spectrum of abdominal procedures. However, liver resections have for long remained resistant to the onslaught of laparoscopic surgery, despite a first report as early as 1992. Concerns regarding the difficult mobilization and transection of the liver and the risks of major haemorrhage, gas embolism and dissemination of malignant tumours have been responsible for this initial slow development. More recently, however, increased experience in laparoscopic cancer surgery and the contribution of improved technology have fuelled the enthusiasm for laparoscopic liver surgery. Increasing numbers of reports and a recent meta-analysis suggested that the laparoscopic approach to the anteroinferior segments of the liver (II–VI) is a suitable alternative to open surgery, being as safe as open surgery. Laparoscopic liver surgery is associated with reduced blood loss, reduced postoperative morbidity and shorter hospital stay, and these findings are now well recognized. This has culminated in the recent acclamation of the laparoscopic approach as a gold standard, at least for selected procedures such as left lateral sectionectomy. Together with a shift in practice towards more limited, anatomic, segmental resections and parenchyma-sparing surgery, laparoscopic liver surgery is now a well-established technique routinely employed in specialized liver surgery units.

GUIDE TO FURTHER READING

Liver function tests, acute and chronic liver disease, congenital and genetic liver disease

Rodes J, Behamou J-P, Biel A, *et al. Textbook of Hepatology*, 3rd edn. London, UK: Blackwell Publishing, 2007.

Anatomy and intraoperative ultrasound

Skandalakis JE, Skandalakis LJ, Skandalakis PN, Mirilas P. Hepatic surgical anatomy. *Surg Clin North Am* 2004;**84**:413–435.

Torzilli G, Makuuchi M. Ultrasound guided finger compression in liver subsegmentectomy for hepatocellular carcinoma. *Surg Endosc* 2004;**18**:136–139.

Portal hypertension and complications

Arroyo V, Colmenero J. Ascites and hepatorenal syndrome in cirrhosis: pathophysiological basis of therapy and current management. *J Hepatol* 2003;**38**(Suppl. 1):S69–89.

Bismuth M, Funakoshi N, Cadranel JF, Blanc P. Hepatic encephalopathy: from pathophysiology to therapeutic management. *Eur J Gastroenterol Hepatol* 2011;**23**:8–22.

Chavez-Tapia NC, Barrientos-Gutierrez T, Tellez-Avila FI, *et al.* Antibiotic prophylaxis for cirrhotic patients with upper gastrointestinal bleeding. *Cochrane Database Syst Rev* 2010;**9**:CD002907.

European Association for the Study of the Liver. EASL clinical practice guidelines on the management of ascites, spontaneous bacterial peritonitis, and hepatorenal syndrome in cirrhosis. *J Hepatol* 2010;**53**:397–417.

Garcia-Tsao G, Sanyal AJ, Grace ND, Carey W. Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. Practice Guidelines Committee of the American Association for the Study of Liver Diseases; Practice Parameters Committee of the American College of Gastroenterology. *Hepatology* 2007;**46**:922–938.

Hashizume M, Akahoshi T, Tomikawa M. Management of gastric varices. *J Gastroenterol Hepatol* 2011;**26**(Suppl. 1):102–108.

Häussinger D, Schliess F. Pathogenetic mechanisms of hepatic encephalopathy. *Gut* 2008;**57**:1156–1165.

Jalan R, John TG, Redhead DN, *et al.* A comparative study of emergency transjugular intrahepatic portosystemic stent-shunt and esophageal transection in the management of uncontrolled variceal hemorrhage. *Am J Gastroenterol* 1995;**90**:1932–1937.

Moore KP, Aithal GP. Guidelines on the management of ascites in cirrhosis. *Gut* 2006;**55**(Suppl. 6):vi1–12.

Runyon BA. AASLD Practice Guidelines Committee. Management of adult patients with ascites due to cirrhosis: an update. *Hepatology* 2009;**49**:2087–2107.

Scottish Intercollegiate Guidelines Network. *Management of Acute Upper and Lower Gastrointestinal* Bleeding. Edinburgh, UK: SIGN, 2008.

Sharma P, Rakela J. Management of pre-liver transplantation patient – part 2. *Liver Transpl* 2005;**11**:249–260.

Tripathi D, Hayes PC. Endoscopic therapy for bleeding gastric varices: to clot or glue? *Gastrointest Endosc* 2008;**68**:883–886.

Wong F, Nadim MK, Kellum JA, *et al.* Working Party proposal for a revised classification system of renal dysfunction in patients with cirrhosis. *Gut* 2011;**60**:702–709.

Liver transplantation

Devlin J, O'Grady J. Indications for referral and assessment in adult liver transplantation: a clinical guideline. British Society of Gastroenterology. *Gut* 1999;**45**(Suppl. 6):VI1–VI22.

Mazzaferro V, Chun YS, Poon RTP, *et al.* Liver transplantation for hepatocellular carcinoma. *Ann Surg Oncol* 2008;**15**:1001–1007.

Murray KF, Carithers RL Jr. AASLD practice guidelines: Evaluation of the patient for liver transplantation. *Hepatology* 2005;*41*:1407–1432.

Neuberger J, Gimson A, Davies M, *et al.*; Liver Advisory Group; UK Blood and Transplant. Selection of patients for liver transplantation and allocation of donated livers in the UK. *Gut* 2008;*57*:252–257.

Liver resection and liver metastases

Adam R, Azoulay R, Castaign D, *et al.* Liver resection as a bridge to transplantation for hepatocellular carcinoma on cirrhosis. A reasonable strategy? *Ann Surg* 2003;**238**:508–519.

Cuschieri A. Laparoscopic liver resection. *J Min Access Surg* 2005;**1**:99–109.

De Jong MC, Van Dam RM, Maas M, *et al.* The liver first approach for synchronous colorectal liver metastases: a 5 year single centre experience. *HPB (Oxford)* 2011;**10**:745–752.

Hillingso JG, Wille-Jorgensen P. Staged or simultaneous resections of synchronous liver metastases from colorectal cancer. A systematic review. *Colorectal Dis* 2009;**11**:3–10.

Liau KH, Blumgart LH, De Matteo RP. Segment-oriented approach to liver resection. *Surg Clin N Am* 2004;**84**:543–561.

Polignano FM, Quyn AJ, de Figuereido RS, *et al.* Laparoscopic versus open liver segmentectomy: prospective, case-matched intention-to-treat analysis of clinical outcomes and cost effectiveness. *Surg Endosc* 2008;**22**:2564–2570.

Poon RTP. Current techniques of liver transection. *HPB (Oxford)* 2007;**9**:166–173.

Poon RTP, Fan ST, Lo CM, *et al.* Long term survival and pattern of recurrence after resection of small hepatocellular carcinoma in patients with preserved liver function. Implications and strategy of salvage transplantation. *Ann Surg* 2002;**3**:373–382.

Terminology Committee of the International Hepato-Pancreato-Biliary Association. The Brisbane 2000 Terminology of Liver Anatomy and Resections. *HPB (Oxford)* 2000;**2**:333–339.

Topal B, Aerts R, Penninckx F. Laparoscopic intrahepatic Glissonian approach for right hepatectomy is safe, simple and reproducible. *Surg Endosc* 2007;**21**:2111.

CHAPTER 25

Disorders of the biliary tract

SIR ALFRED CUSCHIERI

Surgical anatomy

Biliary tract

The right hepatic duct is formed by the intrahepatic union of the dorsocaudal and ventrocranial branches draining the two sectors of the right liver (segments V–VIII) (Figure 25.1). The ventrocranial duct is in direct line with the right hepatic duct and crosses in front of the dorsocaudal branch as this arches downwards before reaching the confluence of the two ducts. The left hepatic duct, which is formed by medial and lateral branches draining segments II–IV, is longer than the right hepatic duct. It follows a partial extrahepatic course (of variable length depending on the width of the quadrate lobe) and, therefore, dilates readily in the presence of distal obstructive disease.

The extrahepatic portion of the left duct and its segment III branch can be accessed surgically at the hilum by following the insertion of the round ligament (ligamentum teres) in the depths of the recessus of Rex (Figure 25.2). This 'round ligament' approach is an effective method of bilioenteric bypass for inoperable cholangiocarcinoma (CC) of the extrahepatic ducts.

The union of the right and left hepatic ducts is usually extrahepatic (90% within 1.0 cm of liver parenchyma), high up in the porta hepatis. The resulting common hepatic duct receives the cystic duct lower down, whereupon it becomes the common bile duct. It is customary, however, in surgical anatomy to use the term 'common bile duct' or simply 'bile duct' for the entire extrahepatic conduit as it obviates difficulties in nomenclature, especially when there is a low insertion of the cystic duct. The junction of the right and left hepatic ducts is also referred to as the *hilar bifurcation*. Together with the hepatic artery to its left and the portal vein behind, the common bile duct is surrounded by fibrous tissue known

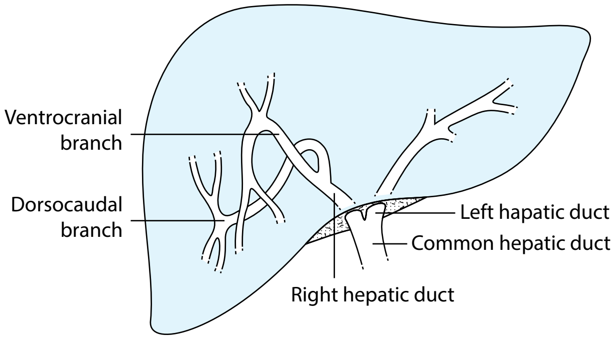


Figure 25.1 Frontal schematic view of the liver illustrating the intrahepatic biliary tree and the ductal arrangement at the hilus. Note the left duct has a longer extrahepatic course. Within the liver, the dorsocaudal branch of the right hepatic duct curves acutely posterior to the ventrocranial branch.

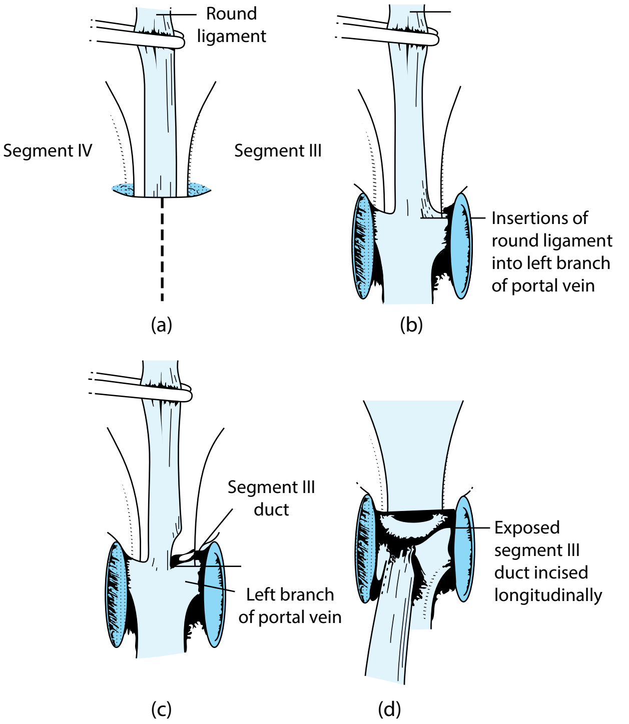

Round ligament
Segment IV
Segment III
Insertions of round ligament into left branch of portal vein
(a)
(b)
Segment III duct
Left branch of portal vein
Exposed segment III duct incised longitudinally
(c)
(d)

Figure 25.2 Diagrammatic representation of the round ligament approach to segment III duct: (a) the round ligament leads to the umbilical fissure between segment IV and segment III which are often joined by a bridge of liver tissue that overlaps the terminal insertions of the round ligament into the left branch of the portal vein; (b) the bridge of liver tissue has been divided to expose the vascular terminations of the round ligament to the left branch of the portal vein; (c) the terminations of the round ligament have been suture ligated and divided to expose the left branch of the portal vein; (d) downward traction displaces the left branch of the portal vein with exposure of the segment III duct which is divided longitudinally for anastomosis to a loop of jejunum.

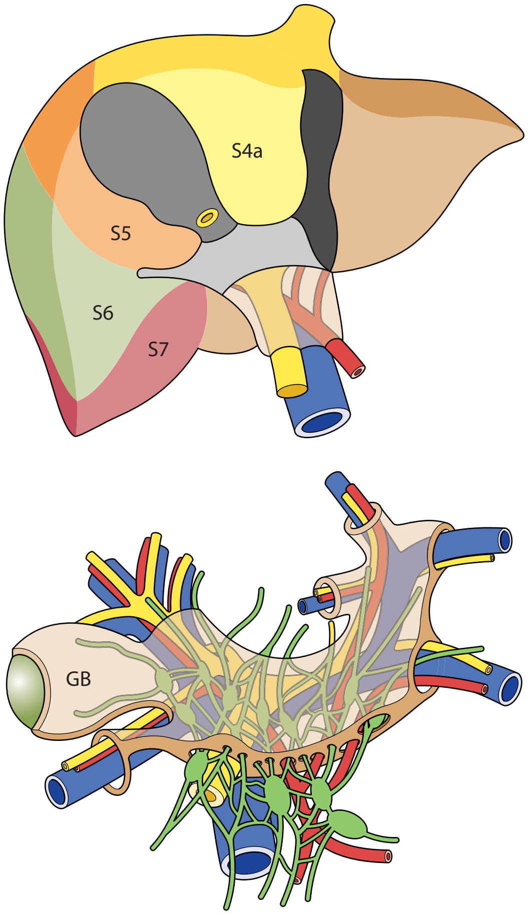


Figure 25.3 Hepatic plate system. (a) Location of the different plates; (b) components of the plate system. GB, gallbladder; G, Glisson's sheath (From Kawarada, *J Hepatobiliary Pancreat Surg* 2000;**7**:580–586, with permission.)

as the Glissonian sheath. At the hilum this is thickened and forms a condensation that is often referred to as the *hilar plate*. If the liver is incised anteriorly and posteriorly (between the hilum and the caudate lobe) to the hilar plate, finger dissection enables the mobilization of the main divisions of the hepatic duct, hepatic artery and portal vein (Figure 25.3). This manoeuvre allows inferior displacement and thus access in case of high bile duct strictures. It is also used for segmental resections of the liver.

Hepatic plate system

The hepatic plate system consists of bile ducts and blood vessels (arteries and veins) surrounded by a fibrous sheath which intrahepatically is continuous with Glisson's capsule and extrahepatically with the hepatoduodenal ligament. In addition, the sheath contains meshwork lymphatic vessels, nerves and a scant vascular network. Whereas the plate system surrounds the bile ducts and hepatic artery, i.e. they are located within the plate system, the portal vein is not and is simply covered by a loose connective tissue sheath. This is the reason for the easy dissection and separation of the extrahepatic bile duct and hepatic artery from the portal vein at the hilum of the liver.

The limits of the hilar plate are *above* by segment 4a (the inferior part of the medial segment), Rouviere sulcus (demarcating the junction between S6 and S5) *on the right* and continuous with the cystic plate (beneath the gallbladder), whereas *on the left* it is continuous with the umbilical plate. As the anterior Glisson's sheath generally runs behind the junction between the cystic plate and the hilar plate, and the posteroinferior Glisson's sheath runs behind the Rouviere sulcus, the bile ducts and blood vessels of the right side can be dissected easily without widely opening the hilar plate. Rouviere's sulcus is a surface fissure between the right lobe and the caudate process present in the vast majority of patients. It corresponds to the level of the porta hepatis where the right pedicle enters the liver.

Ductal anomalies

The intrahepatic arrangement outlined above applies in 75% of cases. A different arrangement is encountered in the remainder when either the right dorsocaudal or ventrocranial ducts join the left hepatic duct, or the common hepatic duct forms a trifurcation (Figure 25.4). The majority (75–80%) of intrahepatic calculi are located in the left hepatic duct, and right-sided calculi, which are far less common, are usually found in the ventrocranial branch of the right hepatic duct.

Important extrahepatic anomalies sometimes referred to as 'aberrant ducts' are encountered in 15–19% of patients. In fact, these 'anomalous/aberrant ducts' represent an extrahepatic confluence of a segmental duct and in the vast majority (95%) affect the right system, when the aberrant duct joins the right hepatic duct (extrahepatically), or common hepatic duct or cystic duct and, very rarely, the gallbladder.

Gallbladder and cystic duct

The gallbladder is a pear-shaped sac about 10 cm in length and is situated on the inferior surface of segment V of the right liver. It is covered with a layer of peritoneum that contains many small veins that require coagulation during cholecystectomy. It is customarily divided into the fundus, which has the poorest blood supply, especially when the organ is distended, the body and the neck or infundibulum, which leads to the cystic duct. Not infrequently, the neck has an abnormal sacculation, which is referred to as Hartmann's pouch. This may become adherent to the surrounding structures of the porta hepatis, particularly the common bile duct, seriously obscuring anatomical relationships during dissection of this region.

The cystic duct runs a variable course from the neck of the gallbladder to join the common hepatic duct. Its mucosa is arranged in a spiral fold or valve (valve of Heister), which often causes difficulties in cannulation during operative transcystic cholangiography. Although most anatomical textbooks indicate that the cystic duct joins the bile duct along its right margin, several large series of surgical dissections and analyses of operative cholangiograms demonstrate clearly that this arrangement is rare and is only encountered in 15–20% of cases. Much more commonly, the cystic duct enters the bile duct either posteriorly or anteriorly (40%). It may also pursue a spiral or a parallel course with the bile duct, with the two structures being enclosed in a common fibrous sheath that tends to obscure the exact location of the entry of the cystic duct into the bile duct (Figure 25.5). The spiral cystic duct runs down and behind the common hepatic duct to enter on its medial aspect (35%). The parallel cystic duct runs parallel to the bile duct for a variable distance before entering it. This is the rarest arrangement and is encountered in 5–7% of patients. Rarely, the cystic duct joins the right hepatic duct and very infrequently the left duct.

Anomalies of the gallbladder

The most common anomaly of the gallbladder encountered during surgery is the Phrygian cap, where the fundus is constricted and turned back on itself. The fully intrahepatic gallbladder is rare. The so-called floating gallbladder, which has a complete serosal covering and a dorsal mesentery, is relatively uncommon, as is malposition of the gallbladder and double gallbladder.

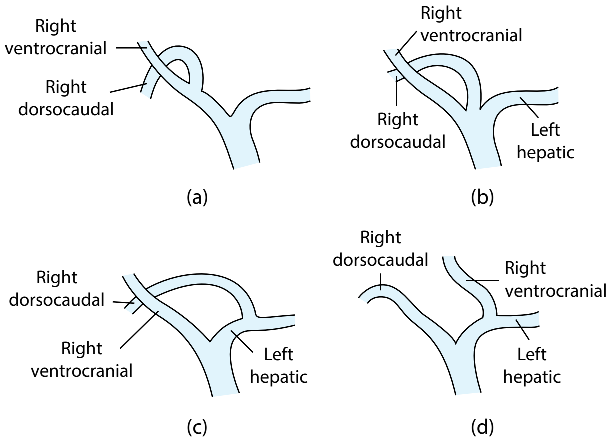


Figure 25.4 Variations in the confluence of the major intrahepatic ducts: (a) normal arrangement which is found in 75% of cases; (b) trifurcation where the right ventrocranial, right dorsocaudal and the left hepatic ducts arise simultaneously from the common hepatic duct, there being no right hepatic duct; (c) termination of the right dorsocaudal branch into the left hepatic duct; (d) termination of the right ventrocranial duct into the left hepatic duct.

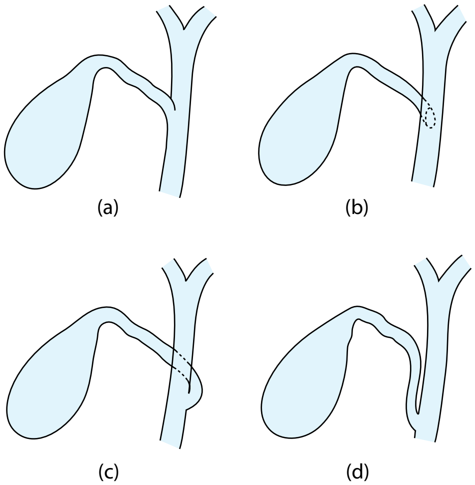


Figure 25.5 Schematic representation of the termination of the cystic duct: (a) lateral insertion often depicted as the usual arrangement but which is only encountered in 15–20% of patients; (b) anterior or posterior termination – this is the most common type and accounts for 40% of cases; (c) spiral cystic duct which courses behind the bile duct to open on its medial aspect – this is fairly common and is found in 35% of patients; (d) parallel cystic duct – this is the rarest arrangement and is encountered in 5–7% of patients.

The floating gallbladder predisposes to torsion, which simulates acute cholecystitis. An elongated sausage-shaped gallbladder frequently accompanies congenital cystic disease of the bile ducts. Agenesis (congenital absence) of the gallbladder is very rare and the condition can only be diagnosed at laparotomy in a patient who has not undergone previous biliary tract surgery. Another rare anomaly is the trabeculated gallbladder, but this usually causes symptoms similar to chronic cholecystitis and is associated with abnormal gallbladder emptying.

A left-sided gallbladder is an integral component of *situs inversus.* In the absence of this condition, malposition is generally regarded as a very rare anomaly. Thus a collective review of the Western literature yielded only 24 cases of left-sided gallbladder. Two types of gallbladder malposition have been described – *medioposition* of the gallbladder and *sinistroposition* (transposition). In medioposition, the gallbladder is displaced medially to lie on the undersurface of the quadrate lobe (segment IV) but still on the right side of the round ligament. In sinistroposition, the gallbladder lies under the left lobe (segment III) to the left of the round ligament (Figure 25.6). Despite its alleged rare incidence, the authors of a joint two-centre (Dundee and Eindhoven) study have encountered five cases of sinistroposition in a consecutive series of 1764 patients undergoing laparoscopic cholecystectomy (LC), a prevalence of 0.28%. The resulting pathological anatomy has implications for the safe conduct of LC. Despite the left-sided location of the gallbladder, the biliary pain experienced by these patients is always on the right side. The preoperative diagnosis of this anomaly is made only rarely despite routine preoperative external ultrasonography and selective recourse to endoscopic retrograde cholangiopancreatography (ERCP).

In sinistroposition, the cystic artery always crosses in front of the common bile duct from right to left. The cystic duct may open on the left or right side of the common hepatic duct. The anomaly does not preclude safe LC but modifications of the port sites and the use of the falciform lift facilitate the procedure in these cases.

The arterial supply of the gallbladder is by means of the cystic artery, which usually arises from the right hepatic artery. The cystic artery is an end artery and its occlusion is followed by gangrene of the gallbladder. There are several congenital anomalies of the arterial supply of the gallbladder (Figure 25.7), the most important of which is a short cystic artery arising from a looped right hepatic artery. All these arterial anomalies are however important and must be recognized during cholecystectomy before ligature of the 'cystic artery'. Careful display and verification of the anatomy is the single most important factor in the prevention of arterial bleeding and iatrogenic injuries during cholecystectomy and biliary tract surgery.

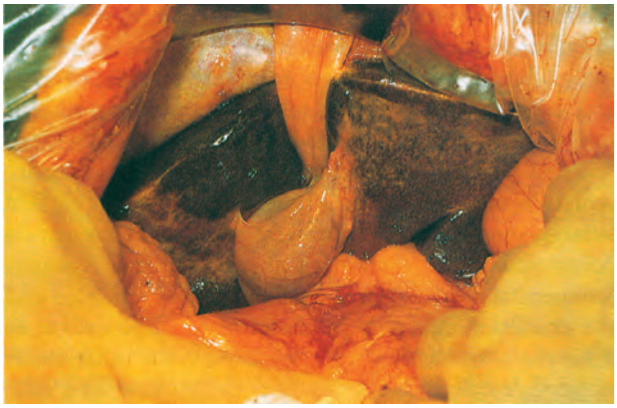


Figure 25.6 Operative photograph of sinistroposition of the gallbladder which lies under the left lobe (segment III) to the left of the round ligament. Despite the left-sided location of the gallbladder, the biliary pain experienced by these patients is always on the right side.

Common bile duct

The bile duct (choledochus) is formed by the union of the right and left hepatic ducts each draining the respective hemiliver. It is joined at a variable distance along its course by the cystic duct. In strict anatomical terms, the segment between the hilar bifurcation and the cystic duct is referred to as the common

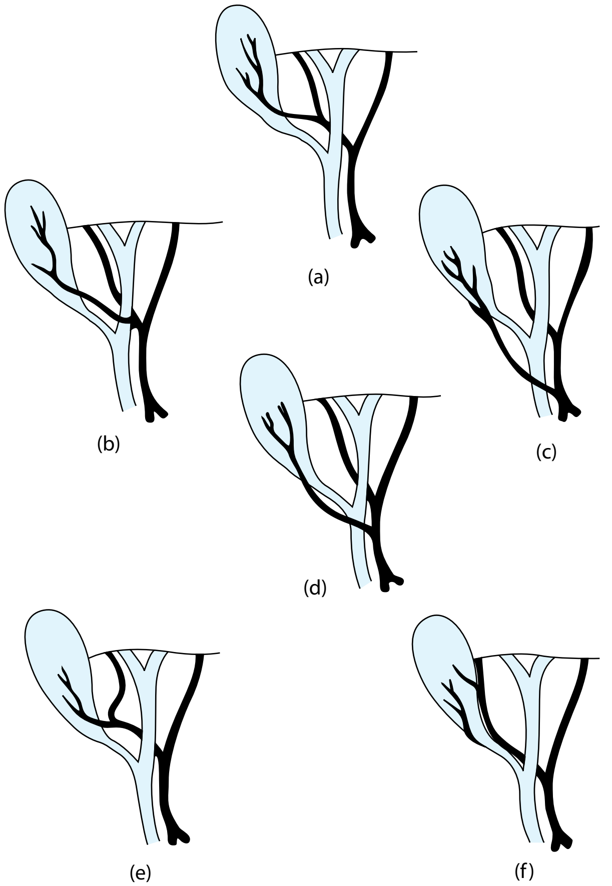


Figure 25.7 Anomalies of the cystic artery: (a) normal arrangement where the cystic artery arises from the right hepatic artery soon after this emerges from behind the common hepatic duct; (b) origin of the cystic artery from the right hepatic to the left of the bile duct, the cystic artery then crossing in front of the common hepatic duct; (c) low origin of the cystic artery from the common hepatic or gastroduodenal arteries; (d) accessory cystic artery arising from the hepatic artery – this second artery can also arise from the left hepatic, right hepatic and gastroduodenal arteries; (e) looped right hepatic artery with a short cystic artery arising from the summit of the right hepatic arterial arch; (f) the right hepatic runs close to the cystic duct and the neck of the gallbladder before giving anterior and posterior cystic branches – this anomaly is the most dangerous since the right hepatic is easily mistaken for a large cystic artery.

hepatic duct and the term common bile duct is reserved for the portion distal to this junction. From the surgical standpoint, however, it is best to consider it as one structure, which is divisible into the *supraduodenal*, *retroduodenal*, *intrapancreatic* and *intraduodenal* segments. It serves as a conduit of bile from the liver and gallbladder to the duodenal papilla and in the adult measures 11–12 cm in length with an average diameter of 7 mm (range 4–10 mm).

The supraduodenal segment is important surgically because it is the area that is most commonly explored. It lies in the free edge of the hepatoduodenal ligament to the right of the hepatic artery and anterolateral to the portal vein. The retroduodenal segment curves to the right away from the portal vein behind the first part of the duodenum before entering the head of the pancreas – intrapancreatic segment. However, in 20% of patients the duct has a partial or complete extrapancreatic course. The transduodenal segment (also known as the infundibulum), which traverses obliquely the duodenal wall and usually joins the pancreatic duct, opens into the duodenal lumen at the summit of the major duodenal papilla. The lower end of the common bile duct, therefore, deviates to the right before entering the lumen of the duodenum almost at right angles. This is an important practical consideration since forcible probing through this area may perforate the bile duct and result in a haematoma, postoperative pancreatitis, choledochoduodenal fistula or stricture of the lower end of the bile duct.

The main pancreatic duct (Wirsung) joins the posteromedial wall of the transduodenal segment of the bile duct to form a common channel in 90% of cases. A localized dilatation of the common channel to form an ampulla of Vater is uncommon (10–20%) and in 10% of patients the two ducts open separately into the duodenum (Figure 25.8).

The Vaterian segment includes the lower 2.5–3.0 cm of the common bile duct, the distal part of the pancreatic duct, the ampulla or common channel and the major duodenal papilla. These structures are surrounded by a condensation of circular and longitudinal smooth muscle fibres often referred to as the sphincter of Oddi, although it was Boyden who described the detailed anatomy of the various components of this sphincteric complex. The inferior sphincter is the strongest component and is also known as the papillary muscular ball (Figure 25.9). It surrounds the terminations of the bile and pancreatic ducts and the common channel. The middle sphincter is the longest and the thinnest of the components and surrounds the transduodenal and a variable portion of the transpancreatic segments of the bile duct and the duct of Wirsung. The superior sphincter consists of localized thickenings of the middle sphincter around the bile and pancreatic ducts at the proximal end of the sphincter complex.

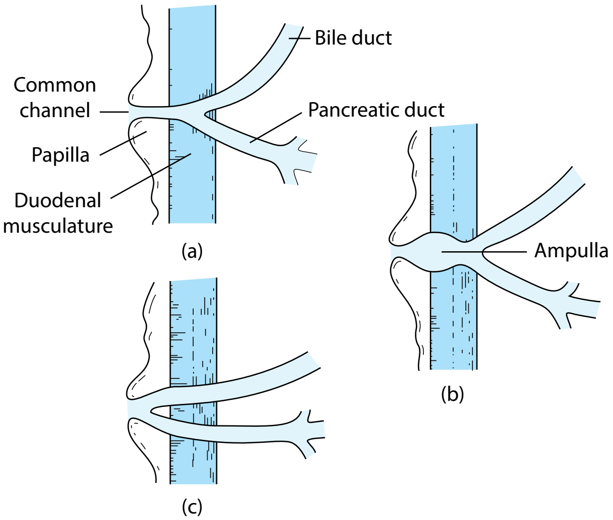


Figure 25.8 Configuration of the lower end of the common bile and pancreatic ducts: (a) the two ducts join to form a common channel, which opens at the summit of the duodenal papilla – this is the most common arrangement; (b) there is a localized dilatation of the common channel to form the ampulla of Vater; (c) the two ducts open separately into the duodenum.

An important variation of the anatomy of the Vaterian segment is the condition known as *pancreas divisum,* which results from failure of fusion of the ventral and dorsal pancreas during embryological development. The duct of the ventral pancreas, which normally forms the main pancreatic duct, remains rudimentary and drains the lower portion of the pancreatic head and the uncinate process. The rest of the pancreas is drained through the duct of the dorsal anlage (duct of Santorini), which opens through the small accessory papilla above the major duodenal ampulla. The incidence of pancreas divisum in the general population is 5–8% but the condition is much commoner in patients with idiopathic recurrent pancreatitis (approximately 25%) and an aetiological relationship has been suggested.

The rest of the common bile duct contains few muscle fibres. Its epithelial lining rests on a loose stroma containing elastic fibres, which disappear with age or disease. Thus, stone impaction, prolonged distension or cholangitis may lead to rigidity of the common bile duct. The narrowest portion of the common bile duct occurs at its point of entrance into the duodenal wall and this area is often indicated by a notch on the cholangiogram. The diameter of the transduodenal segment is normally 5 mm and that of the major duodenal papilla varies from 0.5 to 1.5 mm. The commonest site for calculus arrest or impaction is just proximal to the transduodenal segment. The major duodenal papilla is situated on the posteromedial aspect of the second

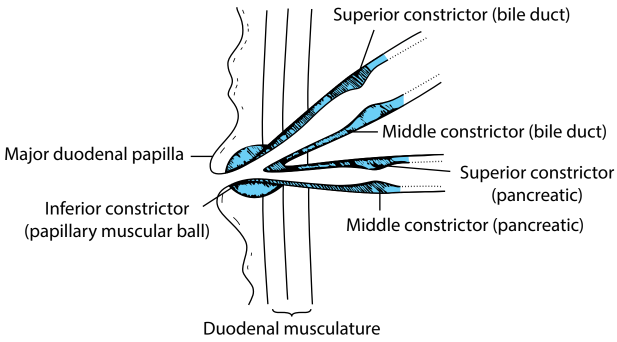

Superior constrictor (bile duct)
Middle constrictor (bile duct)
Major duodenal papilla
Superior constrictor (pancreatic)
Inferior constrictor (papillary muscular ball)
Middle constrictor (pancreatic)
Duodenal musculature

Figure 25.9 Diagrammatic representation of the components of the sphincter complex (sphincter of Oddi) surrounding the Vaterian segment. The bile and pancreatic ducts are illustrated splayed apart to facilitate the demonstration. Normally the terminal portions of both ducts are contiguous.

part of the duodenum about 7.0–10.0 cm from the pylorus. Its appearance may vary from the usual well-defined papilla with varying degrees of projection to a flattened depression between the mucosal folds. Irrespective of its exact configuration, the major duodenal papilla frequently has a dorsal mucosal fold. The papilla is more easily located by ERCP than by direct inspection during surgical intervention. The minor (accessory) papilla is more proximally situated and assumes clinical importance only in patients with pancreas divisum.

The activity of the choledochal sphincteric complex is independent of the duodenal musculature but may be influenced by it. Thus, the effect of certain drugs on the choledochal sphincter differs from their action on the duodenal wall, and duodenal muscular peristaltic activity has no significant effect on the common bile duct pressure. The choledochal sphincter is an active structure and measures up to 2.5 cm in length. It consists of well-developed longitudinal and circular smooth muscle. Contraction of the longitudinal muscle tends to open the duct lumen, whereas the circular muscle has the opposite effect. These contracted (systolic) and relaxed (diastolic) states of the choledochal sphincter lead to quite distinct appearances of the lower end of the common bile duct at cholangiography. During contraction, contrast often forms a meniscus with the concavity facing downwards simulating a stone (the pseudocalculus phenomenon) (Figure 25.10).

Cystohepatic triangle of Calot

The cystohepatic triangle is important in biliary surgery especially in the performance of cholecystectomy. It is a triangular fold of peritoneum containing the cystic duct, cystic artery, cystic node and a variable amount of fat. It also often contains the right hepatic artery, which usually enters the triangle behind the common hepatic duct and before it gives off the cystic artery, after which it curves upwards along the right hepatic duct to the liver. The cystic lymph node is most commonly situated at the junction of the cystic with the common hepatic artery. The vast majority of aberrant/anomalous bile ducts arise from the right ductal system (especially the dorsocaudal branch of the right hepatic) and 80% are located in the cystohepatic triangle of Calot (Figure 25.11). The cystohepatic triangle is virtually obliterated in the presence of Mirizzi syndrome (see below).

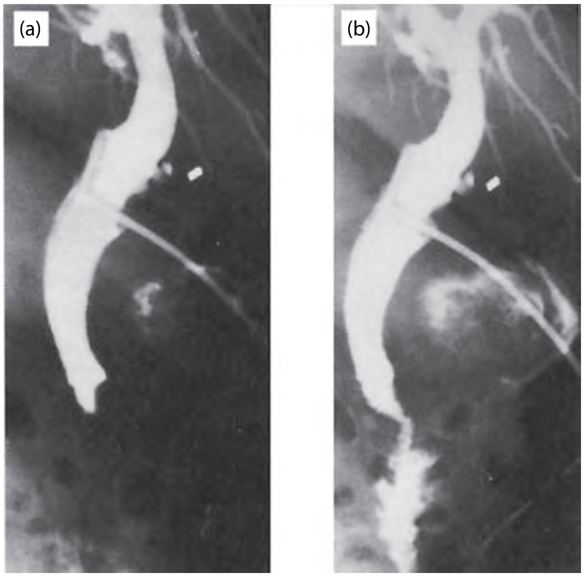


Figure 25.10 The pseudocalculus phenomenon: (a) during the systolic phase of choledochal sphincter action, there is an apparent detachment of the lower end of the common bile duct outline from the duodenal contrast shadow with an inverted meniscus effect, the appearance simulating a stone impacted at this site; (b) this 'filling defect' disappears when the same bile duct is visualized during diastole (relaxation) of the sphincter.

Lymphatic drainage of the gallbladder

Proximally, the lymphatic channels of the gallbladder communicate with those of Glisson's capsule of the liver. The hepatic capsular lymphatics drain into the thoracic duct except those on the superior surface of the liver, flow from which reaches the retrosternal lymph nodes via several channels. Distally, the gallbladder lymphatics and those of the extrahepatic bile ducts drain into the cystic lymph node, which is situated near the cystic artery close to its origin from the right hepatic artery, and to other nodes lateral to the lower end of the bile duct, particularly the retroduodenal segment. The distal lymphatic drainage of the gallbladder follows three routes: (1) cholecystoretropancreatic lymphatics, which run downwards anterior and posterior to the common bile duct to end in the retroportal nodes at the posterior surface of the head of the pancreas; (2) cholecystocoeliac lymphatics, which run to the left through the hepatoduodenal ligament and drain in the coeliac nodes; and (3) the cholecystomesenteric lymphatics, which run to the left in front of the portal vein and drain in lymph nodes at the root of the superior mesenteric vessels. The cholecystoretropancreatic drainage is considered the main pathway with respect to nodal spread from cancer of the gallbladder. The three lymphatic pathways ultimately drain in the abdominal aortic nodes.

Hepatic artery

The adult hepatic artery is derived from the middle of the three primordial arteries that supply the fetal liver. The usual arrangement is for the common hepatic artery to arise from the coeliac axis. After giving off the right gastric and gastroduodenal

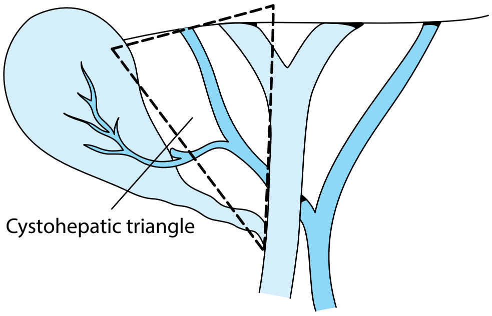


Figure 25.11 Cystohepatic triangle of Calot formed by the cystic duct and neck of the gallbladder inferiorly, the liver edge superiorly and the common hepatic duct medially. It contains the cystic artery and lymph node and the right hepatic artery as it emerges from behind the common hepatic duct. The vast majority of aberrant/anomalous bile ducts arise from the right ductal system (especially the dorsocaudal branch of the right hepatic) and 80% are located in the cystohepatic triangle of Calot.

arteries behind the antroduodenal region, it arches upwards along the left side of the bile duct and in front of the portal vein. It then bifurcates into the right and left hepatic arteries usually quite close to the liver. The right hepatic artery usually crosses behind (rarely in front of) the common hepatic duct before giving rise to the cystic artery. Low division of the hepatic artery is encountered in 15% of patients when the right hepatic artery courses behind the portal vein.

The important anomalies are the result of persistence of the left or right primordial hepatic arteries. The most common (20%) is persistence of the left primordial artery. This anomalous vessel then arises from the left gastric artery or directly from the aorta and traverses the lesser omentum to enter the liver in the umbilical fissure. It is very rare for a persistent left primordial hepatic artery to be the main or only arterial supply to the liver; usually it is present in addition to the normal hepatic artery. Persistence of the right primordial artery results in an anomalous right hepatic artery originating from the superior mesenteric. It ascends to the liver behind the pancreas and duodenum to reach the free edge of the hepatoduodenal ligament. Again it is very rare for a persistent right primordial hepatic artery to provide the sole arterial blood supply to the liver.

Portal vein

The portal vein which measures 8.0 cm in length is formed by junction of the superior mesenteric and splenic (lineal) veins, the union occurring in front of the inferior vena cava behind the neck of the pancreas. The portal vein then courses upwards behind the superior part of the duodenum and ascends to the liver in the right border of the lesser omentum behind the common bile duct on the right and hepatic artery on the left to the right extremity of the porta hepatis, where it bifurcates into the right and left portal veins which accompany the corresponding branches of the hepatic artery into the hepatic parenchyma. The portal vein bifurcation can be intrahepatic or extrahepatic (roughly equal incidence). The portal vein is surrounded by the hepatic plexus of nerves and numerous lymphatic channels and some lymph nodes. The right branch of the portal vein (3.0 cm length) receives the cystic vein before it enters the liver. The left portal vein (3.5 cm length) is of smaller calibre and is joined by the ligamentum teres (obliterated umbilical vein) and may be united by a fibrous cord (ligamentum venosum), the remains of the obliterated ductus venosus. The left portal vein gives branches to the caudate lobe before it enters the liver.

Variant anatomy of portal vein

Many anatomical variants of the portal vein have been described, including duplications, congenital absence and absence of portal vein branching (single portal vein enters the right liver and courses into the left, giving only segmental branches), but all these are rare. An important CT portography-based study of 200 patients has shown that subtle variants are much more common in practice, and these may easily be overlooked with important adverse surgical consequences. In this study, 35% of patients had variant portal vein anatomy (significantly greater than the 10–15% reported in the sonography literature). The findings of this important study are summarized as follows:

- normal bifurcation (standard anatomy) 65%
- trifurcation 9%
- right posterior portal vein as the first branch of the portal vein (Z configuration) 13%
- segment VII as a separate branch of the right portal vein 1%
- segment VI branch as a separate branch of the right portal vein 6%
- other anomalies 6%.

The common abnormal variants affect predominantly the right portal vein, with abnormalities of the left portal vein being rare. Knowledge of abnormal portal vein anatomy is important in surgical resections of the liver, hepatic transplantation and many percutaneous interventional procedures, including transhepatic portal vein embolization to induce contralateral liver hypertrophy in patients, transjugular intrahepatic shunt, etc. If missed, some of these variants may have very serious consequences. Thus for example, failure to recognize a Z-type portal vein variant during a left liver resection or when harvesting a living donor liver transplant may result in loss of perfusion to the right anterior sector. Trifurcation of the portal vein may require two separate anastomoses when the right liver is transplanted.

Surgical biliary physiology

Hepatic cholesterol and bile acid physiology

The liver plays a key role in the metabolism of cholesterol. It regulates the uptake of dietary cholesterol, its *de novo* synthesis and the excretion of this lipid in bile either as free cholesterol or as bile acids. There is increasing evidence for metabolic compartmentalization of cholesterol within the hepatocyte, i.e. cholesterol derived from extrahepatic sources is metabolized differently from that synthesized in the liver, the latter being destined for bile acid synthesis via the 7-α-hydroxylation step, after which it is no longer available for esterification.

Under normal conditions, sufficient bile acids are synthesized to make up for the enteric losses, and to this extent the synthesis of bile acids by the liver is generally believed to be controlled by the circulating bile salt concentration, although there is some debate on the importance of this feedback control mechanism. The primary bile acids are conjugated within the hepatocyte with the amino acids glycine and taurine, before being secreted into the bile canaliculi. Toxic relatively insoluble secondary bile salts (reabsorbed from the small intestinal pool) such as lithocholic acid are sulphated prior to excretion. Cholesterol is transported into the bile canaliculi as phospholipid–cholesterol vesicles. The formation of phospholipid–cholesterol–bile acid micelles is a postcanalicular event induced by the high concentrations of bile acids in the hepatic bile. There is evidence that high concentrations of deoxycholic acid may promote the hepatic secretion of cholesterol saturated bile and thereby induce the nucleation of cholesterol crystals.

Bile secretion

The secretion of bile by the liver is largely dependent on the influx of bile acids, which consist of a large component made up of bile salts that have been reabsorbed from the small intestine into the portal venous blood, and a smaller component of newly synthesized bile acids. There is a small bile acid-independent fraction of bile secretion, which is reduced in experimental cholestasis, whereas the bile acid-dependent fraction increases under these conditions.

Somatostatin decreases both bile acid flow and bile ductular secretion, indicating a suppression of *de novo* bile acid synthesis by the hepatocytes. By contrast, vasoactive intestinal polypeptide (VIP) increases the bile flow in man by a mechanism which is similar to that of secretin. The basal or resting (interdigestive) common bile duct pressure that averages 6.0 mmHg ensures the continued patency of the lumen of the choledochus but is insufficient to overcome the resistance of the choledochal sphincter. The hepatic bile is stored and partitioned in the gallbladder, although the patterns of gallbladder filling and emptying are incompletely understood. Eating results in the cholecystokinetic response, consisting of contraction of the gallbladder and relaxation of the sphincter of Oddi; this results in the timely delivery of bile salts into the intestinal tract. This response is mediated by known hormonal mechanisms [cholecystokinin (CCK)] and poorly understood neural reflexes. Studies have shown that gallbladder emptying occurs before the onset of gastric emptying of a meal, suggesting a cephalic neural reflex. Gallbladder contractions also occur during the interdigestive period, and, in humans, these are coincident with phase II duodenal activity of the intestinal migratory motor complexes.

Biliary motility

The existence of an independent choledochal sphincter is now recognized by electrophysiological criteria. It results in a high-pressure zone just within the papilla. Apart from being responsive to various hormonal influences, the choledochal sphincter has a rich nerve supply via the hepatic vagal plexus, which consists predominantly of motor fibres derived from the left abdominal vagus. The sympathetic component is derived from the spinal segments T8 and 9 via the coeliac and periarterial plexuses. In addition, the lower end of the common bile duct exhibits phasic peristaltic activity with opening (diastolic) and closing (systolic) movements that correlate with pressure waves demonstrated in man. In the fasting state, these phasic contractions are cyclical and parallel the duodenal migratory motor complexes. They are considered to regulate the flow of bile and pancreatic juice into the duodenum. The sphincter is in a contracted state and offers a significant resistance to bile flow in between these phasic contractions.

Effect of hormones

CCK is the main stimulus to gallbladder contraction in response to a meal and CCK administration increases gallbladder pressure, decreases the resistance through the sphincter of Oddi and enhances bile flow in man. CCK-induced gallbladder contraction is largely inhibited by atropine. In humans, the gallbladder is less sensitive to CCK than the exocrine pancreas since the dose of exogenous CCK required to induce trypsin secretion is lower than that necessary for gallbladder contraction. CCK action on the extrahepatic biliary tract is not mediated by the nervous system, although the atropine effect in reducing CCK-mediated contraction suggests that its action may be mediated by the release of acetylcholine from intrinsic cholinergic nerves. CCK administration results in the activation of cyclic adenosine monophosphate. CCK induces relaxation of the sphincter of Oddi, which is concomitant with contraction of the gallbladder, resulting in an efficient delivery of bile into the duodenum in response to a meal. CCK and motilin may be involved with the phasic activity of the choledochal sphincter, which is associated with duodenal migratory myoelectrical complexes during the interdigestive stage.

The activity of CCK depends on the COOH-terminal heptapeptide, a sequence found in gastrin which has a cholecystokinetic activity one-fifteenth that of CCK. Thus, this effect of gastrin is a pharmacological phenomenon in both dog and man, and gastrin has no effect on the biliary pressure of cholecystectomized patients. The synthetic hepta-, octa- and decapeptides of CCK are more potent than the various molecular forms of naturally occurring CCK. In humans, 20 mg/kg of the COOH-terminal octapeptide of CCK (OP-CCK) reduces the gallbladder size by 40%. Caerulein, a peptide of similar composition to CCK extracted from amphibian skin, has a marked cholecystokinetic effect about three times that induced by CCK in the dog. Both the gallbladder contraction and the relaxation of the sphincter of Oddi are produced by a direct action on the smooth muscle. The synthetic derivative, ceruletide, has been used in the relief of biliary colic and to promote passage of retained ductal stones during saline infusion of the common bile duct through the T-tube.

Secretin does not appear to have a significant cholecystokinetic response in humans. Its effects on the gallbladder muscle vary with the species studied. Glucagon relaxes both the gallbladder musculature and the choledochal sphincter in man and dog, probably by direct action. VIP, which has been demonstrated in nerve fibres and neurones within the gallbladder wall, induces relaxation of the gallbladder and inhibits gallbladder contraction induced by CCK. These findings suggest that VIP may act as a local neurotransmitter in the physiological neural regulation of gallbladder function. Somatostatin-containing cells have been demonstrated in the human extrahepatic biliary tract. Somatostatin interferes with gallbladder emptying and reduces bile flow in man. The mammalian tachykinin substance P has been detected in the biliary tract. It is a potent stimulator of gallbladder contraction by direct action and indirectly by inducing the release of a cholinergic transmitter.

Impaired gallbladder emptying

Impaired gallbladder contraction with poor emptying leads to biliary sludge formation, increases the risk of stone formation and in certain susceptible patient groups may lead to acute acalculous cholecystitis. Emptying of the gallbladder is disturbed in:

- patients after vagotomy
- morbidly obese patients
- patients with chronic pancreatitis and exocrine insufficiency
- patients on long-term parenteral nutrition
- diabetics with visceral neuropathy
- critically ill patients.

In the morbidly obese, both fasting gallbladder volume and emptying rates are lower than normal and are associated with precipitation of cholesterol crystals. The cause–effect relationship between these two abnormalities (which precedes which) remains debatable.

One study has shed some light on the mechanism underlying the poor contractile activity of the gallbladder in vagotomized patients. These patients exhibit an abnormally elevated CCK release after a triglyceride meal compared with normal subjects. This has been interpreted as a compensatory response to a reduced sensitivity of the gallbladder to CCK following vagal denervation. In patients with chronic pancreatitis, impaired contractility of the gallbladder is encountered in those patients who develop malabsorption. Impaired gallbladder emptying is thus attributed to reduced fat digestion and inadequate triglyceride absorption leading to defective secretion of CCK. Lack of oral alimentation and abrogation of the postprandial CCK response is involved in impaired motility of the gallbladder encountered in patients on parenteral nutrition.

Delivery of bile into the duodenum

Studies in humans have demonstrated that the bile salt output through an intact choledochal sphincter is significantly lower than that obtained from the T-tube in patients after common bile duct exploration. Whereas peak bile salt output is significantly increased after cholecystectomy, total bile salt output is unaffected by this operation when compared with normal. The exogenous administration of CCK results in the production of a more stable bile secretion than the endogenous release of this hormone induced by the intraduodenal infusion of essential amino acids, and the rhythmic release of bile which follows is related to the concentration of bile salts in the duodenal lumen. The delivery of bile into the duodenum is predominantly influenced by the choledochal sphincter, the activity of which is controlled by complex neurohormonal mechanisms that are as yet incompletely understood. Alcohol ingestion induces spasm of the choledochal sphincter and intraduodenal instillation of 0.1 N HCl produces the same effect in dogs.

Neural influences on biliary motility

In man the results of vagal stimulation by insulin-provoked hypoglycaemia suggest that the parasympathetic nervous system is involved in the maintenance of gallbladder tone. Vagotomy causes dilatation and some delay in gallbladder emptying when studied by cholescintigraphic techniques. Vagotomy also results in a decrease in the nerve fibres within the gallbladder wall. However, the increased prevalence of gallstones after vagotomy suggested by some retrospective reports remains controversial. Vagal activity appears to influence the tone of the choledochal sphincter. Electrical stimulation of the vagus nerve in the dog has been reported to result in no change or in decreased bile flow through the sphincter and an increase in its electromyographic activity. In humans, hepatic plexus vagectomy has been reported to lower the passage pressure (the pressure head which opens the sphincter and permits bile to flow into the duodenum), indicating a lowering of sphincteric muscular activity/contractility.

The role of the sympathetic system remains undefined. An increased threshold for pain induced by biliary distension has been reported in humans following sympathetic blockade. Stimulation of the right splanchnic nerve induces a contraction of the sphincter, which is abolished by α-blockade, and gallbladder dilatation that is inhibited by β-blockade.

Effect of cholecystectomy and sphincterotomy/sphincteroplasty

On *a priori* grounds, removal of the gallbladder would be expected to alter the delivery of bile into the duodenum as this would then depend on the hepatic secretory pressure (maximum 25–30 cmH$_2$O) and the resistance of the choledochal sphincter which is expressed by the passage (yield) pressure. In practice, cholecystectomy in humans does not lead to any alterations in the total bile salt output compared with the normal situation, but there is a redistribution of the pool of bile salts between the gut and the portal venous system. The effect of cholecystectomy on cholesterol saturation of bile is uncertain, with some reports indicating a reduction and others showing no effect on bile composition.

Although no changes in biliary cholesterol content, gallbladder filling and response to CCK are observed after division of the sphincter (endoscopic or surgical), the concentration of lecithin and bile salts in the gallbladder is decreased, as is the concentrating ability of the gallbladder. In addition, the results of animal experiments have shown an increase in monohydroxy and dihydroxy bile salts in the gallbladder, which becomes colonized by bacteria and develops histological changes of chronic inflammation. In humans, cholecystitis develops in 6% of patients with an *in situ* gallbladder within 6 months of an endoscopic sphincterotomy. A further 30% will develop cholecystitis and biliary symptoms during subsequent years.

Investigation of patients with biliary tract disorders

Plain abdominal radiology

Plain abdominal film is of limited use in the diagnosis of gallstones in the elective situation, as only 10% of gallstones are radio-opaque. However, a plain film of the abdomen can provide useful diagnostic information in the acute situation. It may demonstrate gas in the biliary tract in patients with bilioenteric fistulas. The demonstration of this gas together with dilatation of the small intestine provides good evidence of gallstone ileus (intraluminal distal small bowel obstruction by a large gallstone), although the stone itself is rarely visualized on the film. Plain film also provides valuable diagnostic information in patients with emphysematous cholecystitis. Finally, calcification within the gallbladder wall, which is an established risk factor for carcinoma of the gallbladder, is best detected by a plain abdominal film.

Oral cholecystography and intravenous cholangiography

Both techniques can provide adequate visualization of the gallbladder if the serum bilirubin is below 40–50 μmol/L. Oral cholecystography can also demonstrate gallbladder contractility after a fatty meal or after the injection of CCK/ceruletide. Failure of the gallbladder to opacify is followed by repeated investigation using a double-dose cholecystogram. Non-visualization by this method is indicative of a diseased gallbladder if ingestion and absorption of the oral contrast agents (Telepaque) can be reasonably assumed (probability exceeds 90%). The sensitivity of a technically satisfactory oral cholecystogram for the detection of radiolucent stones exceeds 90%. However, visualization of the ducts is poor and is obtained in only 20% of patients after a fatty meal, although CCK or ceruletide cholecystography enhances the visualization of the ducts considerably (80% of patients). In practice, oral cholecystography is very rarely used in the diagnosis of gallstones nowadays and has been replaced by external ultrasound scanning. Oral cholecystography may be useful in the diagnosis of polypoid lesions of the gallbladder.

Likewise intravenous cholangiogram is very rarely used for outlining the extrahepatic biliary tract. Although infusion cholangiography is an accurate technique for the diagnosis of cystic duct obstruction (acute cholecystitis, where the gallbladder is not opacified but the ducts are outlined), this technique has been largely replaced by biliary scintigraphy, which is more accurate. In addition, there is a small but definite incidence of severe hypersensitivity reactions. In practice, preoperative intravenous cholangiogram has been abandoned in most centres.

External abdominal ultrasonography

This is now the first-line investigation for biliary tract and pancreatic disease in most hospitals. Aside from its non-invasive nature and lack of any radiation exposure, ultrasound scanning can provide simultaneous information on the following:

- detection of gallstones
- detection of gallbladder disease and its complications
- dilatation of the biliary tract and hepatic parenchymal disease, e.g. tumour deposits
- detection of neoplasms and disorders of the biliary tract, liver and pancreas.

Detection of gallstones

Real-time ultrasound in experienced hands can detect gallstones in over 90% of cases (Figure 25.12).

Gallstones are detected by ultrasound as gravity-dependent, mobile, echogenic foci within the gallbladder lumen which cast a posterior acoustic shadow. Despite the reported high diagnostic yield, gallstones that are too small (<1 mm), soft stones which lack strong internal echoes and gallstones impacted in the gallbladder neck or in the cystic duct may be missed by external abdominal ultrasound scanning. A special technique which employs harmonic imaging has been shown to improve detection of small gallstones. With this technique, the insonating ultrasound beam is at the fundamental frequency (2.5 or 3 MHz), but the receiving (returning) echoes are both at the fundamental frequency and at the second harmonic frequency (twice the fundamental frequency), enabling the imaging software to significantly reduce degradation of the image by noise. Hence, harmonic imaging, by increasing the echogenicity of gallstones and strengthening their posterior acoustic shadows, permits visualization of small stones. However some stones, e.g. soft brown pigment stones, are often not visible even by harmonic imaging. These stones, because they lack acoustic shadowing, are often misinterpreted as sludge or debris.

Detection of ductal calculi

The sensitivity of external ultrasound for ductal calculi is much less and varies considerably from centre to centre. Ductal calculi are easy to visualize by external ultrasound only when surrounded by anechoic bile in a dilated duct and when they are sufficiently large to cast an acoustic shadow. Stones impacted in the distal duct are thus often difficult to detect as their margins are inseparable from the adjacent ductal walls and surrounding echogenic fat. Ductal calculi smaller than 5 mm are missed unless special techniques such as harmonic imaging are used. With harmonic imaging

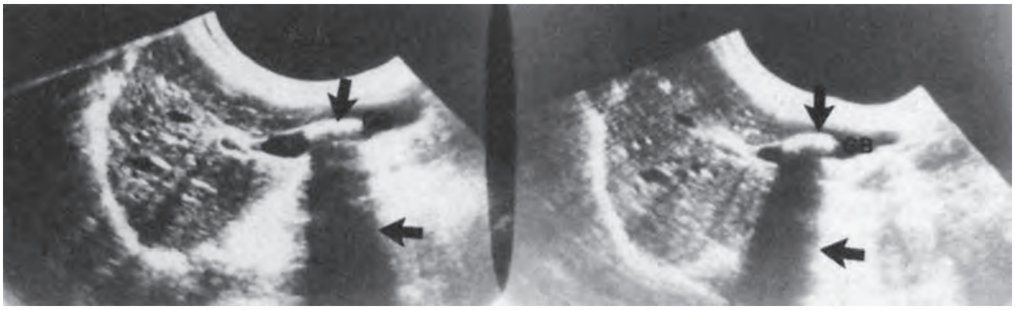


Figure 25.12 Ultrasound examination demonstrating a large gallstone (vertical arrow) with the associated acoustic shadows (transverse arrow) which are characteristic of gallstones. GB, gallbladder.

and related special techniques, the overall reported sensitivity of ultrasound for the detection of ductal calculi is 75%. Difficulties are still encountered in patients without dilatation. In this respect, one-third of common bile duct calculi are found in patients with non-dilated bile ducts. These patients require magnetic resonance cholangiopancreatography (MRCP) or ERCP.

Diagnosis of acute cholecystitis

The diagnosis of acute cholecystitis is established by the presence of a positive sonographic Murphy's sign in patients with documented gallstones by external abdominal ultrasound. Secondary ultrasound findings which support the diagnosis of acute cholecystitis include gallbladder wall thickening (>3 mm), a distended gallbladder and presence of pericholecystic fluid. The sonographic Murphy's sign is defined as a point of reproducible tenderness over the gallbladder elicited by application of pressure by the transducer head. A sonographic Murphy's sign is reported to be 87–92% accurate for the diagnosis of acute cholecystitis. However, the sonographic Murphy's sign may be absent in patients with denervated gallbladders (diabetic patients) and in patients with gangrenous cholecystitis.

Ultrasound examination of the gallbladder is used as the initial diagnostic procedure for acute cholecystitis since it enables the determination of tenderness over the sonographically identified gallbladder and is able to detect pericholecystic fluid collections and gallbladder wall oedema/thickening (ultrasonographic signs of cholecystitis), in addition to sludge and stones. Nowadays, the sensitivity and specificity of ultrasound in the diagnosis of acute cholecystitis approaches that for gallbladder scintiscanning, which is thus less often used for this purpose.

Gallbladder wall thickening and pericholecystic fluid

Gallbladder wall thickening is defined as a wall diameter greater than 3 mm and is present as a non-specific finding in 50% of patients with acute cholecystitis. Other causes of thickening of the gallbladder wall include adenomyomatosis and cancer of the gallbladder. A thickened gallbladder with a characteristic striated appearance (alternating hyper- and hypoechoic layers) in patients with acute cholecystitis is strongly associated with gangrenous cholecystitis.

Pericholecystic fluid is also a non-specific finding as it can be secondary to localized inflammation from any cause. However, type II pericholecystic collection, which is irregular in shape with thick walls, septations or internal debris, is more likely to be associated with gallbladder perforation and abscess formation.

Benign tumours of the gallbladder

Adenomas appear as well-demarcated, polypoid lesions, and are usually less than 2.0 cm in size and are usually solitary. They are rare compared with cholesterol polyps, the polyps encountered in adenomyomatous hyperplasia, which are usually multiple. On ultrasound, gallbladder adenomas are typically echogenic when small, but become more heterogeneous as they enlarge. The premalignant potential of gallbladder adenomas is believed to be low, although this remains uncertain. On ultrasound imaging, it is impossible to identify coexistent dysplasia or carcinoma *in situ* in an adenoma. The current surgical consensus recommends excision of gallbladder polyps greater than 1 cm in patients older than 50 years or any polyp that is growing on sequential ultrasound scans, even if less than 1 cm.

Malignant neoplasms of the gallbladder

Gallbladder carcinoma accounts for 98% of all gallbladder malignancies; the rest include non-epithelial tumours of muscles and nerves, metastases and lymphomas. The major risk factor is chronic cholecystitis, which leads to dysplasia and carcinoma *in situ*. When gallbladder carcinoma is diagnosed by ultrasound, it is usually advanced. The most common imaging presentation of the gallbladder cancer by ultrasound is that of a mass completely replacing the gallbladder (50%). Less commonly focal or diffuse wall thickening (20–30%) or an intraluminal polypoid mass (15–25%) may be seen. In practice, any solitary mass greater than 1.0 cm with internal vascularity should raise suspicion of gallbladder carcinoma. Marked wall thickening greater than 1.0 cm or loss of the normal mural layers are also indicative of malignancy.

Detection of dilated biliary tract

The ultrasonographic detection of a dilated biliary tract is the first step in the investigation of patients with biochemical evidence of cholestatic jaundice. In icteric patients, its accuracy in the diagnosis of extrahepatic bile duct obstruction exceeds 90% (Figure 25.13).

The internal diameter of the common duct is measured (from inner wall to wall) at the level of the crossing of the right hepatic artery. Normally on ultrasound, at this level the diameter should not exceed 6 mm. The diameter of the common bile duct increases slightly to 7–8 mm distally as it approaches the pancreas. There is still some debate whether the bile duct dilates with age and after cholecystectomy. Most centres consider a common bile duct of less than 6 mm to be normal, and one which is greater than or equal to 8 mm as abnormal. Intrahepatic bile ducts are considered normal if they measure 2 mm or less at the porta. With modern ultrasound scanners and use of harmonic imaging technique, intrahepatic bile ducts can be visualized in normal patients.

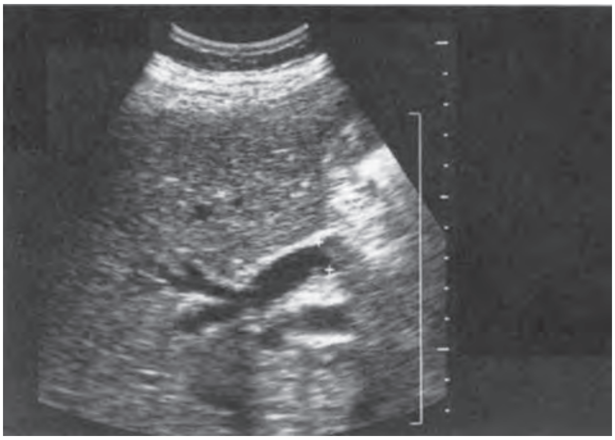


Figure 25.13 Ultrasound examination of the liver showing dilated intrahepatic and common bile ducts in a patient with carcinoma of the pancreas.

Intrahepatic biliary duct dilatation is also accompanied by irregular angular branching, a central stellate configuration, and acoustic enhancement posterior to the ducts. Power Doppler is also extremely valuable in confirming that the dilated intrahepatic tubular structures are indeed bile ducts and not vascular structures. The diagnosis of biliary obstruction by ultrasound requires to be accompanied by clinical signs of obstruction, including elevated bilirubin or elevated alkaline phosphatase. With modern techniques the level of obstruction can be defined in up to 90% of patients and the cause in up to 70%.

Nonetheless, ultrasound examination does not always provide accurate information on the exact site and extent of the lesion causing the extrahepatic obstruction. For this reason, further investigation with CT and MRCP or ERCP is required in most patients. Percutaneous fine-needle cytological aspiration of mass lesions in the liver, extrahepatic bile ducts, gallbladder and pancreas is now routinely performed in many centres under ultrasonographic guidance and may establish the definitive diagnosis regarding the benign or malignant nature of the lesion. There are no known biological hazards of ultrasound investigation (at the energy level used for diagnostic purposes).

Ultrasound examination may prove unsatisfactory for technical reasons in the following:

- in obese patients
- after previous surgery
- in those with ascites
- in those with gaseous distension of the upper abdominal viscera.

In these instances, CT scanning provides more reliable information.

CT scanning

Helical multidetector CT scanning can provide similar information on the biliary tree as ultrasonography, but in view of cost and radiation exposure it is usually held in reserve when ultrasound examination provides insufficient information for any reason. Contrast-enhanced CT provides much better detection of solid lesions of the liver, extrahepatic bile ducts (e.g. CC) and pancreas.

The contraindications to CT include pregnancy (absolute) and history of allergy/hypersensitivity to iodinated intravenous contrast medium (applies only to contrast-enhanced CT).

Two techniques of contrast-enhanced CT are used:

- *Single-phase (portal phase) contrast-enhanced CT* – used in the majority of routine liver CT scanning when the liver is imaged during the peak of parenchymal enhancement (when contrast medium-laden portal venous blood is perfusing the liver). This occurs at about 60–70 seconds after the bolus injection. The slice thickness used depends upon the CT scanner specification and is usually 5 mm or less.
- *Multiphase contrast-enhanced CT* – multiple scanning of the liver after a single bolus injection of contrast. Most liver tumours receive their blood supply from the hepatic artery, in contrast to normal hepatic parenchyma, which receives 80% of its blood supply from the portal vein. Hence hepatic tumours will be strongly enhanced during the arterial phase (at 20–25 seconds after bolus injection) but will have the same density as the normal parenchyma during the portal venous phase. Early and late arterial phase with portal venous phase is used in patients with suspected hepatocellular cancer.

Magnetic resonance imaging

MRI, also known as nuclear magnetic resonance, is being increasingly used for the detection of hepatobiliary and pancreatic disease at the expense of CT, especially with the advent of 3.0T machines, which provide extremely high resolution. The technique is based on the behaviour of protons (e.g. hydrogen) in the nuclei of molecules, which act as spinning magnets and align themselves in a specific direction when exposed to an external electromagnetic field. If radiofrequency (magnetic) pulses corresponding to the spinning frequency of the protons are then applied, the alignment is disturbed with each pulse and then returns or 'relaxes' to the original position. The time taken for the nuclear motion to get out of step is known as T_2 and that required for the return to the original position is known as T_1. Both these time constants vary with the proton density of the tissue and the local atomic and molecular environment of the protons (i.e. chemical state of the tissue). The movement of the nuclei resulting from the externally applied radiofrequency pulses is known as 'resonant absorption' and is accompanied by the re-emission of radiowaves that are picked up by a receiver coil placed round the patient. Different techniques of radiofrequency pulses are used to generate images of different slice thickness and in various planes. The advantages of MRI are:

- excellent soft-tissue definition, particularly of the central nervous system because of the lipid/water content
- it can image in any plane (coronal, sagittal, transverse) without movement of the patient
- it avoids radiation exposure
- it carries no known biological hazard.

It is likely that magnetic resonance spectroscopy will be able to provide the best test of liver function in the near future. Already, it has been used successfully to identify specific inborn errors of metabolism due to specific enzyme deficiencies.

The production of multiplanar images of the pathological anatomy by MRI accounts for its usefulness in the detection of the site and cause of obstruction in patients with cholestatic jaundice. MRI (if available) should be used in preference to CT in patients with suspected hilar CC and primary carcinoma of the gallbladder, where it is superior to CT in assessing the presence and extent of extramural invasion.

Magnetic resonance cholangiopancreatography

The principle of MRCP is based on heavily weighted T_2 pulse sequences (obtained by fast spin echo sequences) since these ensure a very high signal from stationary liquids such as bile and pancreatic juice because of their long T_2 relaxation time, in marked contrast to the surrounding vascular liver parenchyma with its high blood flow that generates virtually no signal because of its much shorter T_2. Hence the bile column stands out as a hyperintense signal against the background hypointense liver tissue. As a result, detailed imaging of the entire biliary tree (intra- and extrahepatic) and pancreatic ductal system is possible without the administration of contrast in the vast majority (>90%) of patients.

Studies have shown that MRCP provides useful clinical information that influences management in the following:

- *Biliary calculi* – MRCP detects stones as small as 2 mm in the bile duct even when this is not dilated. With the right sequences it has a sensitivity of 90–95% and specificity of 90–100% (overall diagnostic accuracy of 95%) for the detection of choledocholithiasis. This has been shown to approximate to the diagnostic accuracy of ERCP (Figure 25.14). Pitfalls for stone detection by MRCP include surgical clips, pneumobilia, haemobilia/clots and flow artefacts. In particular, surgical clips used to secure the medial end of the cystic duct after cholecystectomy create signal void artefacts that obscure the related segment of the common bile duct. MRCP should displace ERCP for the detection of ductal calculi in patients with gallstone-associated acute pancreatitis, and hence the need when these are obstructing the papilla for endoscopic sphincterotomy. MRCP is the best technique for demonstrating the presence and exact location of intrahepatic stones. MRCP is also very useful in differentiating common hepatic duct obstruction by Mirizzi syndrome from obstruction by gallbladder cancer or enlarged lymph nodes.
- *Choledochal cysts* – MRCP displays all types of choledochal cysts.
- *Bile duct injury* – MRCP outlines both the ducts and the perihepatic bile collection but is unable to establish whether the bile leak is active or not. Thus in this situation both MRCP and ERCP are needed.
- *Bile duct tumours* – MRCP has distinct advantages over ERCP and transhepatic cholangiography in the diagnosis of high bile duct tumours because it provides anatomical information of the entire biliary tract on both sides of the stricture. In addition it detects extension of tumour along intrahepatic ducts and enables complete tumour staging of the disease, i.e. assesses involvement of liver, portal nodes and portal vein (with additional MR pulse sequences, e.g. contrast-enhanced T_1 weighted spin echo). MRCP is also valuable in assessing the patency of hepaticojejunostomy following surgery for high bile duct strictures or hilar cholangiocarcionoma (Figure 25.15).

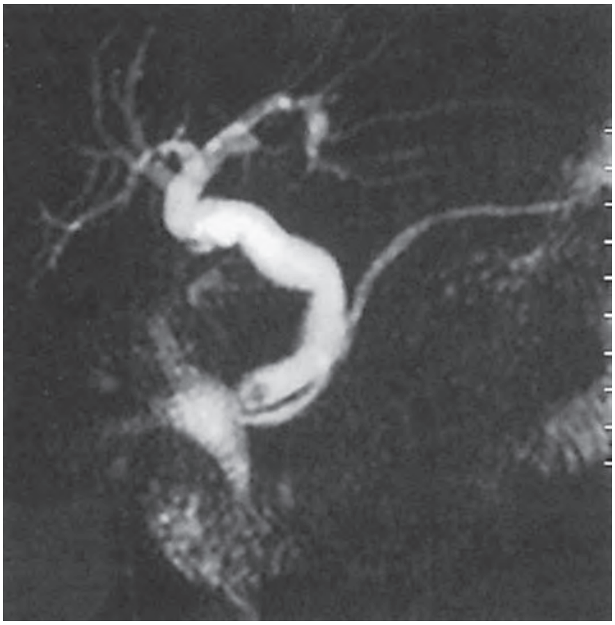


Figure 25.14 Magnetic resonance cholangiopancreatography outlining the entire biliary tree and demonstrating a stone at the lower end of the bile duct. Note also that the pelvicalyceal system of the right kidney is visualized and this obscures the duodenum in this patient who was admitted with acute pancreatitis. The entire pancreatic duct is clearly visible.

- *Pancreatic disease* – the main pancreatic duct is visualized in over 95% of cases. It is highly accurate in the diagnosis of pancreas divisum. In established chronic pancreatitis, MRCP has the same diagnostic accuracy as ERCP but not in early disease with minimal changes in the side branches. MRCP is very useful for ampullary tumours – it detects a dilated bile duct abutting on an irregular ampullary mass indenting the duodenum. Although it can detect proximal and body cancer of the pancreas, it offers no advantages over CT/ultrasound in this respect, except in cystic tumours in view of its ability to detect static fluid collections.

MRCP vs ERCP

MRCP gives the same diagnostic information as ERCP except for the detection of minimal disease chronic pancreatitis. Its distinct advantage is its entirely non-invasive nature. In comparative studies, the failure rate of MRCP (contraindications to MRI, claustrophobia and inadequate examination) is lower than ERCP. MRCP is applicable irrespective of altered or pathological anatomy that precludes ERCP, e.g. duodenal stenosis, hepaticojejunostomy. Thus the trend is to replace diagnostic ERCP with MRCP and restrict ERCP to those patients who require endoscopic intervention, i.e. endoscopic sphincterotomy and stone extraction, stenting, etc.

Percutaneous transhepatic cholangiography

Nowadays, this is a less commonly used technique for the visualization of the biliary tract in the jaundiced patient but it

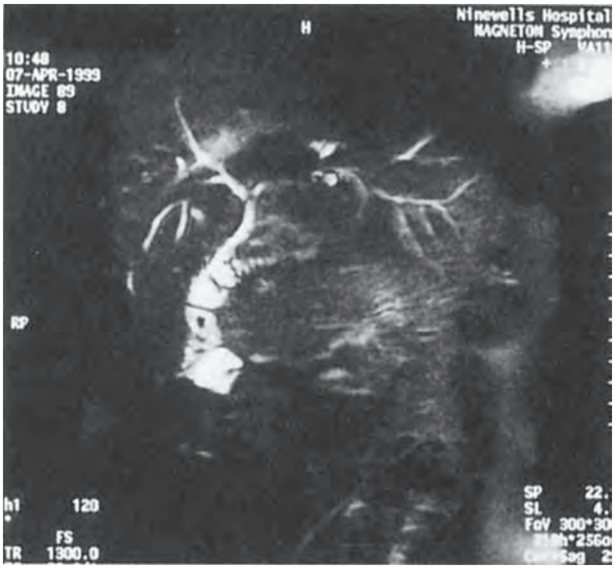


Figure 25.15 Magnetic resonance cholangiopancreatography outlining a hepaticojejunostomy after repair of an iatrogenic high bile duct stricture.

can be modified to allow percutaneous transhepatic drainage and insertion of endoprostheses. In experienced hands, the success rate with percutaneous transhepatic cholangiography (PTC) approximates to 100% in patients with dilated biliary tracts and exceeds 70% in the absence of bile duct dilatation. The reported accuracy of PTC in detecting the level and cause of the biliary obstruction averages 90% (Figure 25.16). Nowadays, the procedure is carried out under sedation using the Chiba (also known as 'skinny') 22-gauge needle, which has an external diameter of 0.7 mm. The use of the Chiba needle has largely replaced the thicker long-dwell trocar cannulas because of the higher success rate and a lower incidence of complications. The procedure must be covered with systemic antibiotic therapy (usually an amino glycoside or cephalosporin or pipericillin) and any clotting abnormality must be corrected with vitamin K and/or the administration of fresh frozen plasma prior to its performance.

In practice, PTC is used when ERCP fails or does not provide sufficient information on the proximal intrahepatic biliary tree because of an obstructing lesion of the extrahepatic bile ducts. However, MRCP can provide this information, and if available should be used in preference to PTC.

Preoperative percutaneous external transhepatic biliary drainage (Figure 25.17) is seldom performed nowadays since

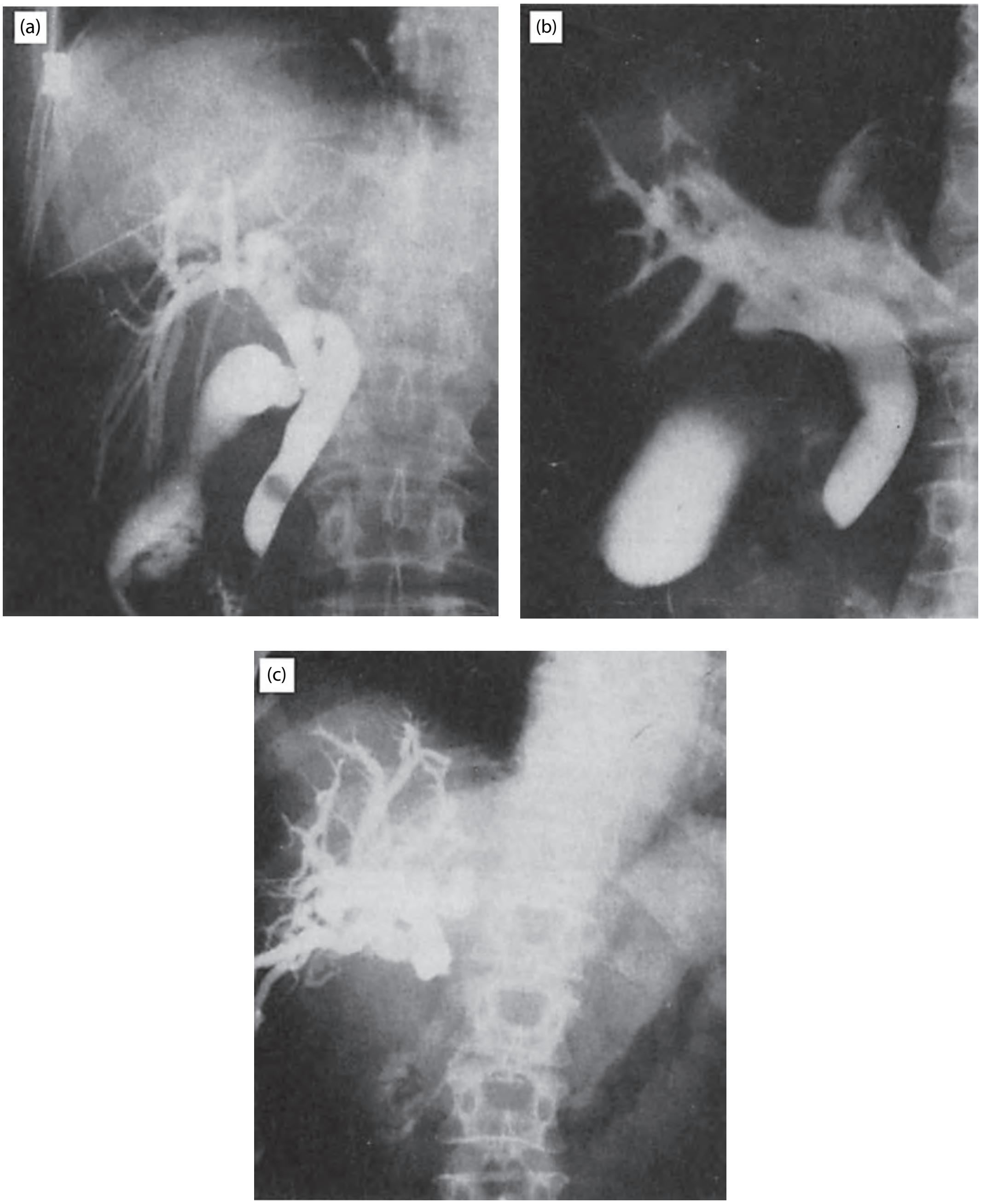


Figure 25.16 Percutaneous transhepatic cholangiograms: (a) ductal calculus proximal to a benign stricture; (b) carcinoma of the pancreas; (c) cholangiocarcinoma.

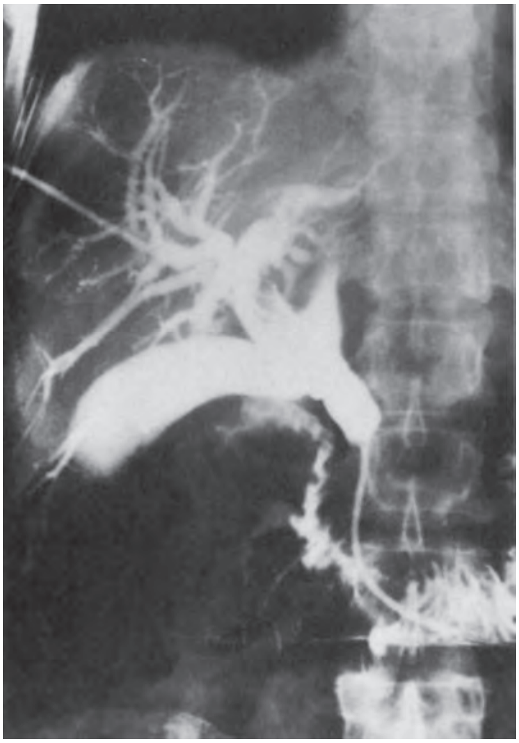


Figure 25.17 Preoperative external biliary decompression for complete obstruction due to carcinoma of the pancreas. This technique is seldom used nowadays except as a prelude to insertion of an endoprosthesis for palliation of complete obstruction caused by inoperable pancreaticobiliary malignancy.

several clinical trials have shown no benefit from the procedure in terms of reduced operative morbidity and mortality in severely jaundiced patients, and the technique predisposes to infection of the obstructed biliary tract, unless a closed collecting system which incorporates bacterial filters is used. However, the percutaneous or endoscopic insertion of endoprostheses (indwelling stents introduced over guide wires and positioned through the obstruction by means of pusher tubes) is a valuable method of palliation of patients with large bile duct obstruction due to inoperable/incurable malignancy (Figure 25.18).

The complications of PTC include:

- bacteraemia
- bile leakage
- haemorrhage: free bleeding into the peritoneal cavity and haemobilia
- bile embolization
- intrahepatic arterioportal fistula
- pneumothorax
- contrast reactions.

The reported incidence of major complications with the Chiba needle varies from 3% to 10% with a mortality of 0.1–0.3%.

Endoscopy and endoscopic retrograde cholangiopancreatography

Upper gastrointestinal endoscopy with a forward- or oblique-viewing panendoscope should be performed in jaundiced patients as significant gastrointestinal pathology is encountered in 25% of jaundiced patients. ERCP, which is performed through a side-viewing endoscope, provides useful information in patients with cholestatic jaundice irrespective of whether the ductal system is dilated or not. In experienced hands, successful cholangiography is achieved by ERCP in over 90% of cases. ERCP permits concomitant endoscopic examination and biopsy of lesions encountered during the endoscopic examination, although the examination of the stomach and duodenum is more difficult and less optimal than with a forward- or oblique-viewing endoscope. A pancreatogram can be obtained during the same investigation. Certain lesions can be treated or palliated during the procedure,

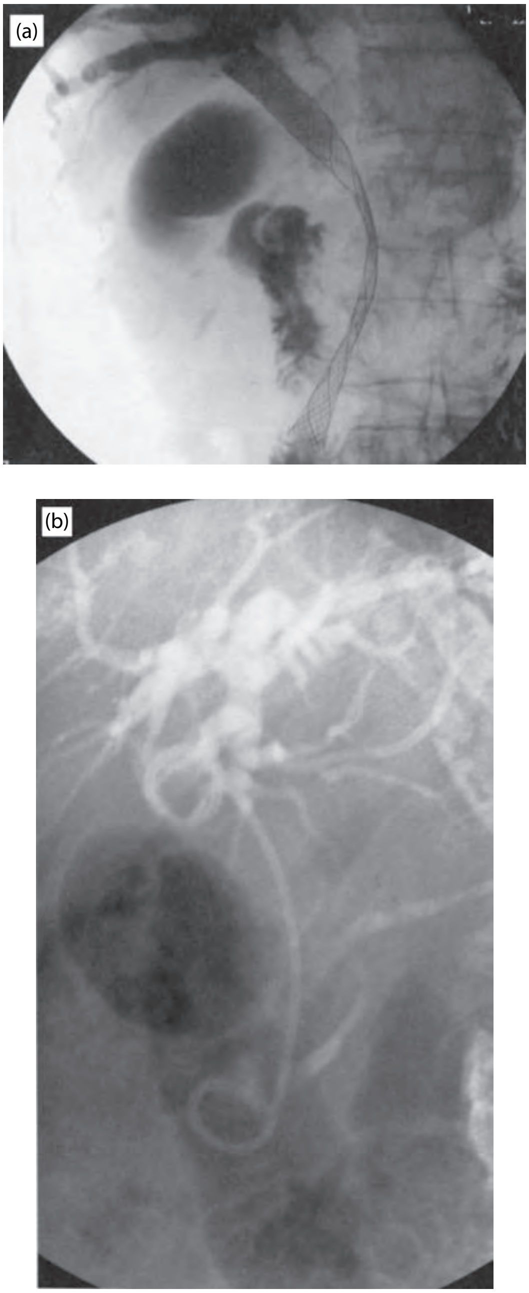


Figure 25.18 Stenting for inoperable disease: (a) expandable metallic endoprosthesis inserted transhepatically under radiological control for palliation of an inoperable hilar cholangiocarcinoma; (b) endoscopic plastic stent for palliation of inoperable carcinoma of the head of the pancreas.

e.g. endoscopic stone removal, endoscopic nasobiliary drainage and stent insertion for inoperable malignant large bile duct obstruction.

Diagnostic ERCP has a very low morbidity, largely due to pancreatitis (1.0%) and mortality (0.1%). The morbidity of interventional (therapeutic) ERCP, especially sphincterotomy, is, however, higher (6–10%). The immediate mortality of endoscopic sphincterotomy averages 1.0%, although the 30 day mortality is 3%. The complications which may follow endoscopic sphincterotomy are:

- haemorrhage
- acute pancreatitis
- cholangitis
- retroperitoneal duodenal perforation
- impacted Dormia basket
- acute cholecystitis
- gallstone ileus – following extraction of large stones.

ERCP is very accurate in the diagnosis of ductal calculi (Figure 25.19), tumours of the bile ducts (Figure 25.20) and pancreas (Figures 25.21 and 25.22) and sclerosing cholangitis. In patients with complete biliary obstructive lesions, the proximal biliary tree is not visualized. These patients require further investigation with MRCP or PTC. ERCP is less accurate than ultrasound and oral cholecystography in the diagnosis of gallbladder disease and gallstones.

Technical failure of an attempted ERCP examination may be due to:

- duodenal or pyloric stenosis
- previous Billroth II gastrectomy
- duodenal diverticulum
- uncooperative patient
- inexperience with the procedure.

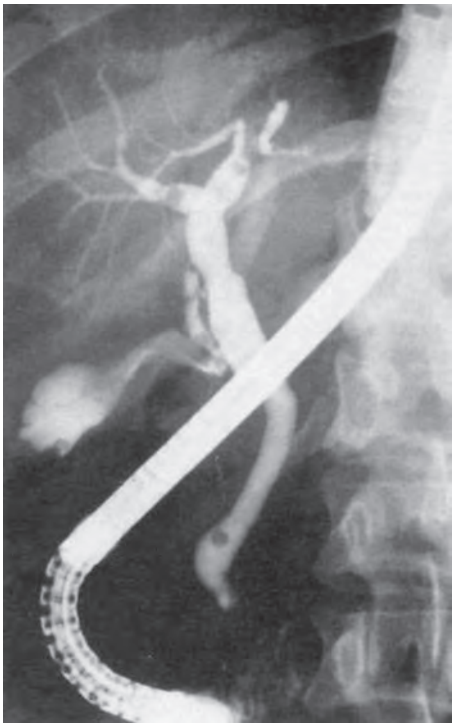


Figure 25.19 Endoscopic retrograde cholangiopancreatography outlining ductal calculi.

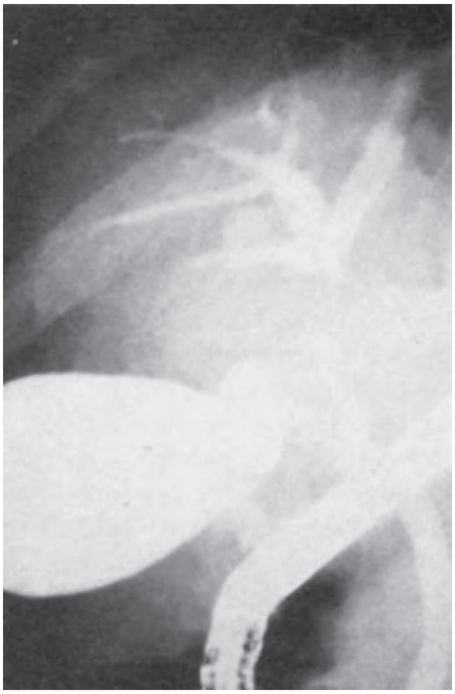


Figure 25.20 Endoscopic retrograde cholangiopancreatography showing a hilar cholangiocarcinoma.

Biliary scintiscanning

The most widely used radiopharmaceutical compounds are ^{99m}Tc-labelled compounds of IDA (iminodiacetic acid), EHIDA (diethylacetanilido-iminodiacetic acid) and more recently ^{99m}Tc-trimethylbromo-iminodiacetic acid (TBIDA). These ^{99m}Tc-labelled IDA derivatives are rapidly cleared from the circulation by hepatocytes and secreted into bile in a similar way to bilirubin; hence their use for the assessment of biliary drainage and gallbladder function. Because TBIDA has high hepatic uptake and low urinary excretion, it gives better visualization of the biliary tract at high bilirubin levels. These agents, which are powerful gamma emitters, are administered intravenously, whereupon they are selectively taken up by the hepatocytes and secreted into the bile. They are therefore ideal for the imaging of the biliary tree by a gamma camera, especially since their uptake by the liver and excretion into the biliary tract is not influenced by the presence of cholestasis.

Biliary scintiscanning is used clinically for the following conditions:

- suspected acute cholecystitis
- assessment of gallbladder, common bile duct and sphincter of Oddi function
- assessment of neonatal jaundice where biliary atresia (BA) is considered
- suspected bile leaks after trauma or surgery
- investigation of biliary drainage.

Biliary cholescintiscanning is the most accurate test of acute cholecystitis irrespective of its nature (acute calculus

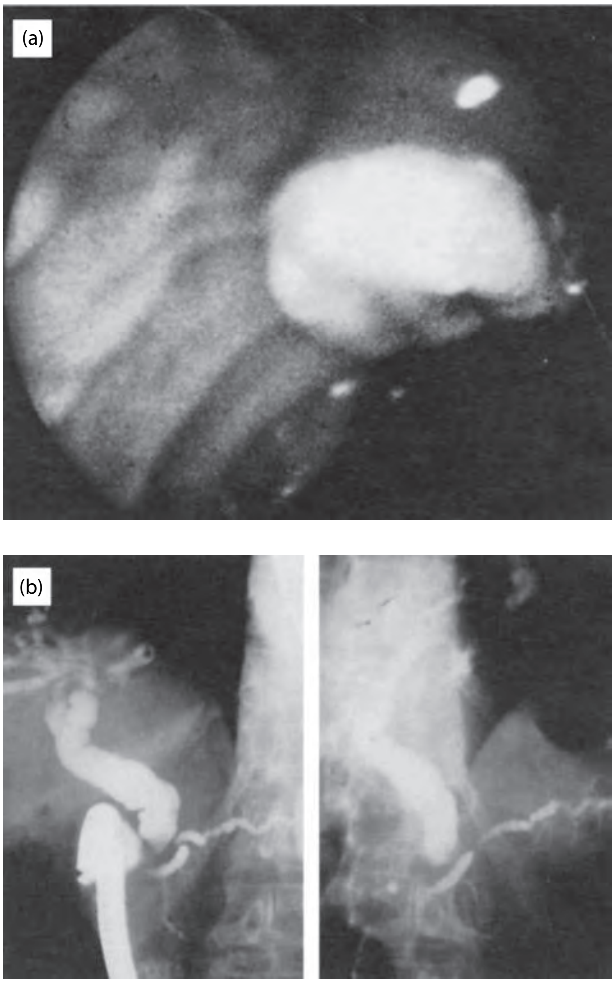


Figure 25.21 (a) Endoscopic view of a periampullary lesion of the pancreas; (b) cholangiopancreatogram of the same patient.

obstructive, acalculous cholecystitis) and establishes the diagnosis within 1 hour of the intravenous administration of the radiopharmaceutical agent. A diagnosis of acute cholecystitis can be confidently made if the scintigram shows prompt excretion and a normal common bile duct with entry of isotope into the duodenum, but the gallbladder is not imaged (Figure 25.23).

The information is stored on magnetic tape/disk for more detailed computer analysis at a later stage. Cholescintiscanning has a sensitivity of 91–97% and a specificity of 87% for the diagnosis of acute cholecystitis. A normal gallbladder scintiscan is virtually 100% accurate in excluding cholecystitis. False positives may be encountered in:

- chronic cholecystitis
- gallstone pancreatitis
- patients with alcoholic liver disease
- patients receiving parenteral nutrition.

The number of false-positive results obtained by cholescintigraphy in the diagnosis of acute cholecystitis can be drastically reduced by the administration of intravenous morphine before the procedure. This opiate causes spasm of the sphincter of Oddi, and thereby induces reflux of bile (and radionuclide) in the gallbladder if the cystic duct is patent. The morphine radionuclide test appears to be very useful in the evaluation of critically ill patients with suspected acute acalculous cholecystitis, in patients who are fasting or receiving parenteral nutrition and as a repeat procedure in those patients in whom the gallbladder was not visualized.

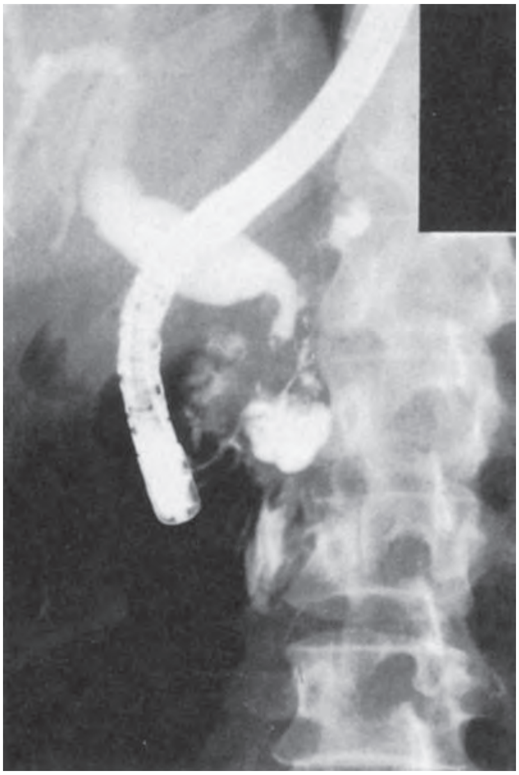


Figure 25.22 Endoscopic retrograde cholangiopancreatography showing a carcinoma of the head of the pancreas.

Biliary scintiscanning has also been used to evaluate the jaundiced patient with a bilirubin greater than 50 μmol/L. Hepatocellular disease is diagnosed when poor liver excretion and intestinal activity are demonstrated after 18 hours of injection. Complete biliary obstruction is denoted by the absence of any intestinal activity after 18 hours, and partial obstruction by normal liver excretion, dilated ducts and delayed intestinal activity. However, biliary scintigraphy is not used routinely for the investigation of the jaundiced patient since other techniques (e.g. ultrasound and cholangiography) give more precise information of the underlying pathology. The exception to the above is provided by jaundice in the neonate where biliary scintigraphy and estimation of faecal radioactivity following the intravenous injection of the isotope is one of the routine tests used for the diagnosis of BA.

Biliary scintigraphy is also very useful for the functional evaluation of surgically constructed bilioenteric anastomoses (Figure 25.24). Biliary scintiscanning after gallbladder contraction induced by a milk meal or intravenous CCK is used to quantify enterogastric reflux. Normal individuals reflux less than 5–10% of the administered dose of the radionuclide. Biliary scintigraphy is also useful in documenting the presence and location of biliary leaks after cholecystectomy. Qualitative analyses based on calculated liver function parameters from scintigraphy dynamic

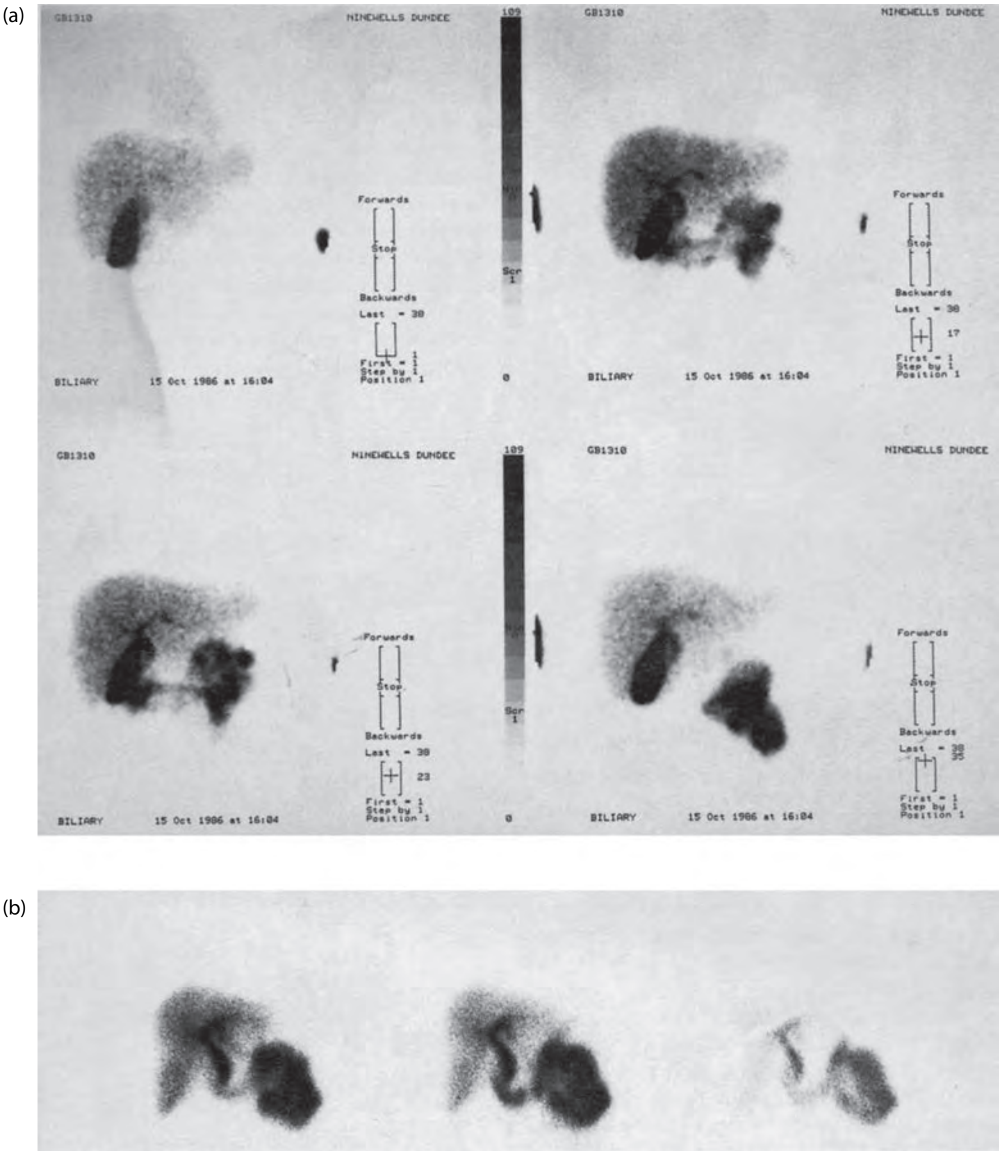


Figure 25.23 Biliary scintigraphy with EHIDA: (a) normal cholescintigram – both the gallbladder and the bile duct are imaged; (b) acute cholecystitis – a normal bile duct with prompt excretion into the duodenum is observed but the gallbladder is not outlined.

studies are used in some centres in the attempt to differentiate between transplant rejection and hepatocyte dysfunction. There are no definite contraindications to biliary scintigraphy.

Cholecystokinin and morphine provocation tests

These involve the use of CCK and/or morphine in combination with TBIDA scanning to improve diagnosis of diseases affecting the gallbladder, common bile duct or sphincter of Oddi. In the CCK test an intravenous infusion of CCK (which contracts the gallbladder) is given over 2–3 minutes and the gallbladder is scanned 30–45 minutes after TBIDA administration. The dynamic imaging is then continued for a further 30 minutes. The tests enable quantitative measurement of the gallbladder ejection fraction and emptying rate. The morphine test is indicated when despite clinical signs of cholecystitis the initial scintiscan is negative or equivocal. Morphine, which contracts the sphincter of Oddi, is administered intravenously (0.04 mg/kg over 1 minute) with scanning for a further 30–40 minutes. Continued non-visualization of the gallbladder for up to 90 minutes confirms the diagnosis of acute cholecystitis. Morphine provocation is also used in diagnosis of motility disorders of the sphincter of Oddi.

Intraoperative fluorocholangiography

Intraoperative fluorocholangiography (IOFC) is commonly performed via the cystic duct (transcystic) but other techniques (direct puncture of the common bile duct, cholecystocholangiography, intraoperative transhepatic cholangiography) are available and used in certain situations.

Transcystic duct fluorocholangiography

Although there are many who practice and advocate the selective use of IOFC during cholecystectomy, in the best

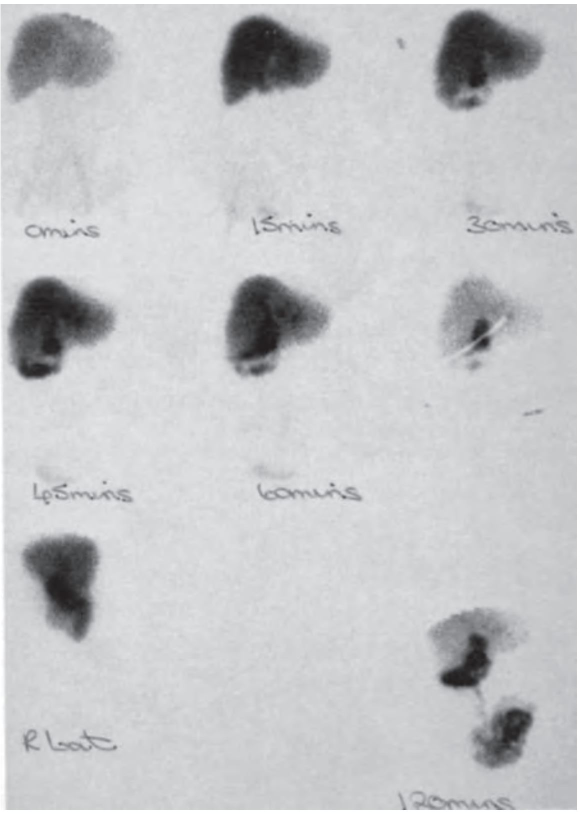


Figure 25.24 EHIDA scintigram showing a functioning hepaticojejunostomy performed for a high bile duct stricture.

interest of the patient this investigation must be regarded as an integral part of cholecystectomy whether this is performed by the open or laparoscopic approach. It provides the best road map of the biliary tree and indicates the need, or otherwise, for exploration of the common bile duct. Although it may not reduce the incidence of bile duct injury during LC, there is good evidence that routine IOFC results in the detection of bile duct damage during the operation. This is important, as the majority of reported bile duct injuries (60%) are missed and discovered in the postoperative period or subsequent to discharge. The arguments for a selective policy for IOFC include increased operating time, unsatisfactory exposures and the safe prediction of a ‘normal’ common bile duct in the absence of a history of jaundice, normal preoperative liver function tests (LFTs) and a normal-sized duct at operation. Intraoperative cholangiography incurs considerable delays and is often unsatisfactory only when performed by a mobile X-ray machine with three blind static films after the sequential injections of contrast medium. The modern procedure entails the employment of portable high-definition C-arm image intensifiers (Figure 25.25).

These have digital facilities and expanded software that enables image storage and advanced image processing: zoom facility, real-time subtraction, road mapping, etc. The entire biliary tree is visualized by screening (fluoroscopy) during injection of contrast medium in all patients inside 5 minutes. Desired images are selected as they appear on the screen and stored on hard disk from which permanent copies can be obtained on X-ray film cassettes.

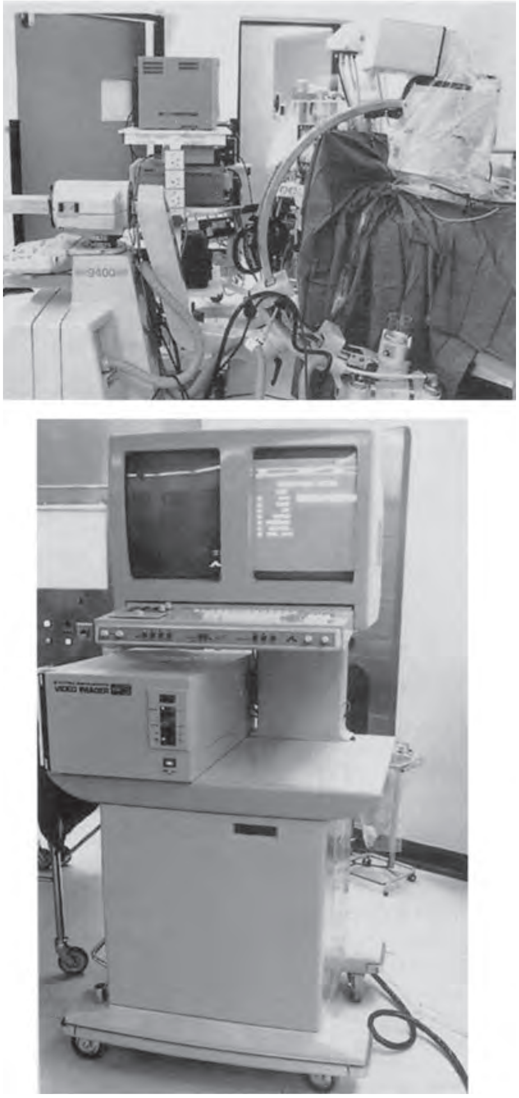


Figure 25.25 Modern C-arm fluoroscopy unit for intraoperative cholangiography.

Unsuspected stones (in patients with normal LFTs and ultrasound examination) are found by digitized fluorocholangiography in normal-sized ducts in 4–7% of patients undergoing cholecystectomy. These stones would all be missed if a selective policy for IOFC were adopted, and although some may pass spontaneously without any adverse effects others will cause acute pancreatitis or cholangitis. Thus the contention that missed stones during cholecystectomy do not matter as they can be treated by endoscopic sphincterotomy when they become symptomatic or cause complications incurs a small but definite morbidity. Anomalies of the biliary tract are encountered in some 19% of patients undergoing routine IOFC during cholecystectomy. Perhaps the most important of these is an abnormal short cystic duct which terminates in either the common hepatic or the right hepatic duct. This anomaly if detected will reduce the risk of bile duct damage, especially during LC where tenting of the extrahepatic conduit is produced as a result of the lateral and upward displacement of the gallbladder. If unrecognized, this will result in partial or total compromise of the common hepatic or common duct by the clip used to secure the medial end of the cystic duct. Finally, a selective policy for IOFC will not impart the experience needed to carry out the investigation expeditiously and the familiarity to interpret accurately the cholangiography findings.

During LC, the most commonly used technique involves cannulation of the cystic duct after this is opened by fine-curved microscissors. Although a variety of disposable purpose-designed catheter systems are available, the best and most cost-effective is the Cook ureter catheter (5 Fr) inserted inside a cholangiograsper (Figure 25.26). This instrument not only guides the catheter to the cystic duct lumen but its jaws allow the cystic duct walls to be grasped on the catheter once this is safely in the lumen of the cystic duct.

Before screening is commenced, the operating table should be tilted 15° to the right (to obviate an intervening spinal column) and in a slight Trendelenburg position to facilitate proximal filling of the ductal system. Nowadays, water-soluble contrast media, such as sodium diatrizoate (25–50%), are used. The volume required depends on dilatation or otherwise of the biliary tract. If during screening a hold-up is encountered at the lower end of the bile duct which does not appear to be organic, intravenous injection of secretin, ceruletide or glucagon is administered when pharmacological relief of spasm of the choledochal sphincter is followed by entry of dye into the duodenum. Contrast injection should be stopped when pancreatic duct filling is observed because of the risk of pancreatitis.

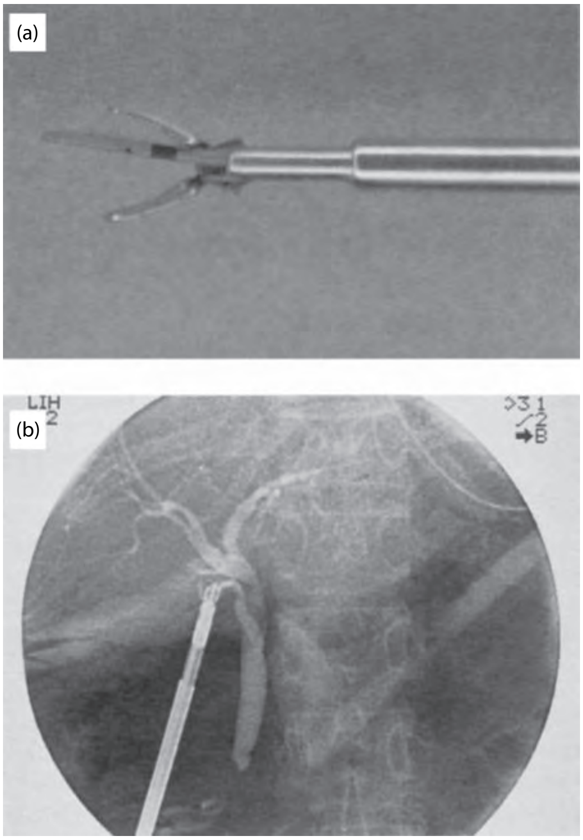


Figure 25.26 Laparoscopic transcystic cholangiography: (a) ureteric catheter inside a cholangiograsper is used for intraoperative transcystic cholangiography during laparoscopic cholecystectomy (LC); (b) intraoperative transcystic cholangiogram during LC using a 4 Fr ureteric catheter and cholangiograsper.

With experience, use of digitized fluoroscopic screening and meticulous technique, both false-negative and false-positive rates of IOFC are very low. Avoidance of metal clips, whenever possible, is desirable as they may obscure small filling defects. The most common artefact leading to unnecessary exploration of the common bile duct remains the air bubble. This can be prevented by ensuring that the delivery system is air free. If 'bubbles' are observed during screening, aspiration should be avoided as this introduces air into the bile duct from the duodenum. In this situation, the syringe should be disconnected from the cannula or catheter, the external end of which is held low, by the side of the drapings, such that the hydrostatic pressure of bile will cause retrograde flow along the cannula and the 'bubble shadow' disappears. Small blood clots, the tips of the transverse processes of the second lumbar vertebra, calcified costal cartilage and a contracted choledochal sphincter with its meniscus effect can also be misinterpreted as ductal calculi when the X-ray machine and static films are used. These sources of error are dismissed easily when fluoroscopic screening is used.

The important rules governing the performance and interpretation of IOFC are listed below.

- Rapid and overfilling of the ductal system must be avoided as, aside from obscuring small ductal calculi, the resulting raised pressure may cause cholangiovenous reflux.
- Unequivocal flow into the duodenum must be demonstrated in all cases.
- Both intra- and extrahepatic bile ducts must be visualized. Non-filling of the intrahepatic biliary tract, or part of it, is always pathological and cannot be ignored. Aside from technical errors, the most common cause for failure of proximal duct filling is a hilar CC. During LC, it may be indicative of common duct injury.
- Contrast injection is stopped with the onset of pancreatic duct filling.
- If doubt exists regarding the interpretation of any abnormalities, expert radiological advice should be sought (Figure 25.27). Ideally, a radiologist should see and comment on the films or the fluoroscopic examination (on hard disk or videotape) routinely.

A completion T-tube cholangiogram is performed after common bile duct exploration, although this is replaced by completion choledochoscopy in some centres. Before injecting contrast, the T-tube and the extrahepatic bile duct is filled with about 60 mL of saline to remove air bubbles. Spasm of the choledochal sphincter is common after biliary manipulations during common bile duct exploration (e.g. passage of balloon catheters). This may result in a hold-up of contrast in the lower end of the bile duct. It is easily differentiated from missed organic obstruction by the intravenous administration of spasmolytic agents such as glucagon or ceruletide (Figure 25.28).

Complications of IOFC are rarely encountered and are usually due to either hypersensitivity to the contrast material or an excessive injection pressure. The latter can result in cholangiovenous reflux and bacteraemia especially in patients with cholangitis.

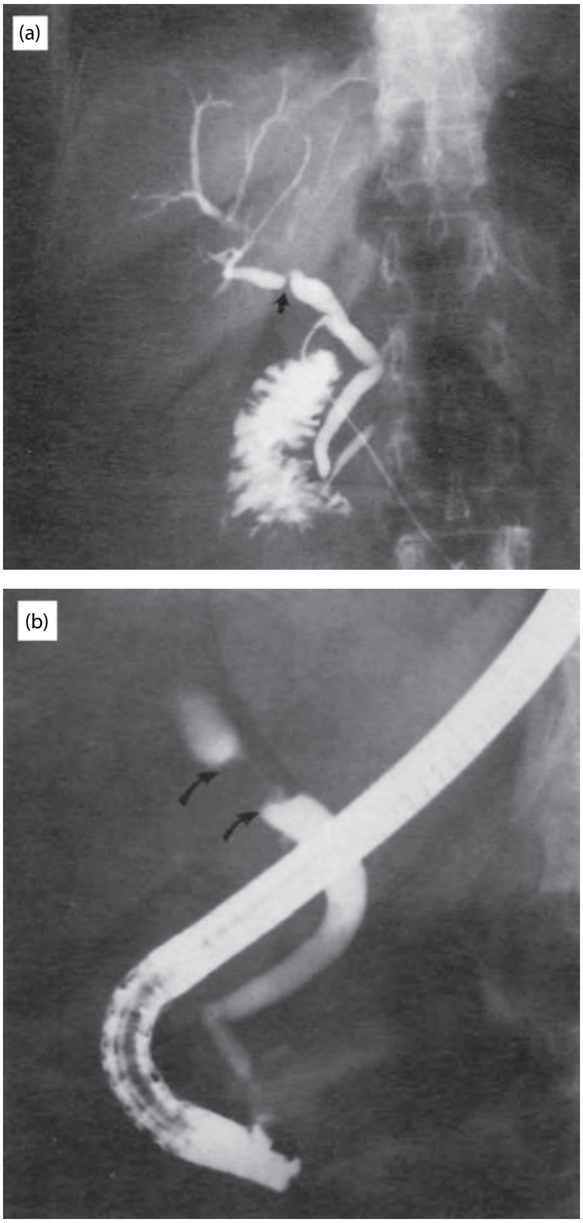


Figure 25.27 (a) Narrowing (arrowed) in the common hepatic duct was thought to be the result of an impression caused by the right hepatic artery. Aside from the obvious stricture, the left intrahepatic ductal system is not visualized. Regrettably, this was overlooked. (b) The patient was referred to the author's unit 3 months later with deep jaundice. The endoscopic retrograde cholangiopancreatography showed an undoubted large carcinoma of the common hepatic duct (outlined between arrows) which proved inoperable.

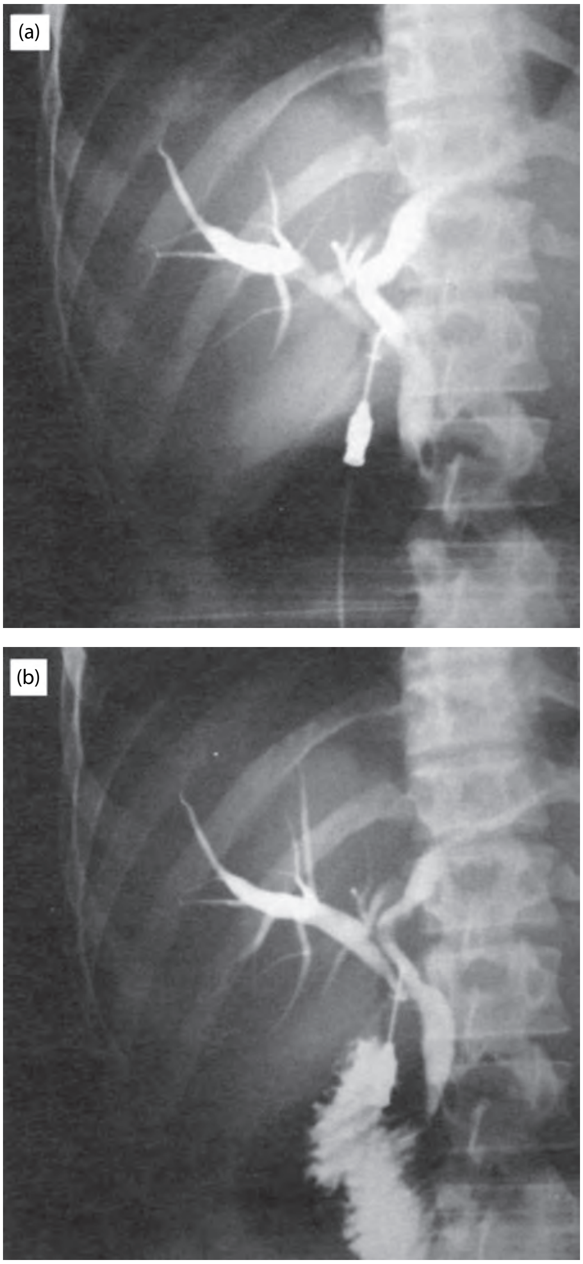


Figure 25.28 (a) Completion intraoperative cholangiogram – there is an apparent hold-up at the lower end which could be due to either spasm or a missed calculus; (b) repeat cholangiogram in the same patient after administration of 1.0 μg of ceruletide as an antispasmodic intravenously. There is now free flow of contrast medium into the duodenum. Alternatively glucagon may be used.

Other methods of intraoperative fluorocholangiograpy

Cholecystocholangiography

This is a simple technique that is primarily indicated when the surgeon is uncertain that the gallbladder needs removal but requires anatomical information of the extrahepatic biliary tract during the operation. Although it has been advocated as an easier alternative to transcystic fluorocholangiography during LC, this practice is ill advised, since forceful injection of large amounts of contrast in the gallbladder may dislodge small stones down a wide cystic duct into the common bile duct. During LC the technique is reserved for when dissection of the triangle of Calot proves difficult.

A cholecystocholangiogram is also indicated if a cholecystojejunostomy is being considered as a palliative procedure in patients with inoperable carcinoma of the pancreas/duodenum with severe obstructive jaundice and itch. The cholecystocholangiogram is performed to determine the patency of the cystic duct and the distance from its entry into the common bile duct and the upper limit of the tumour (Figure 25.29). Only if this distance is greater than 1.5 cm will cholecystojejunostomy suffice as an adequate palliation

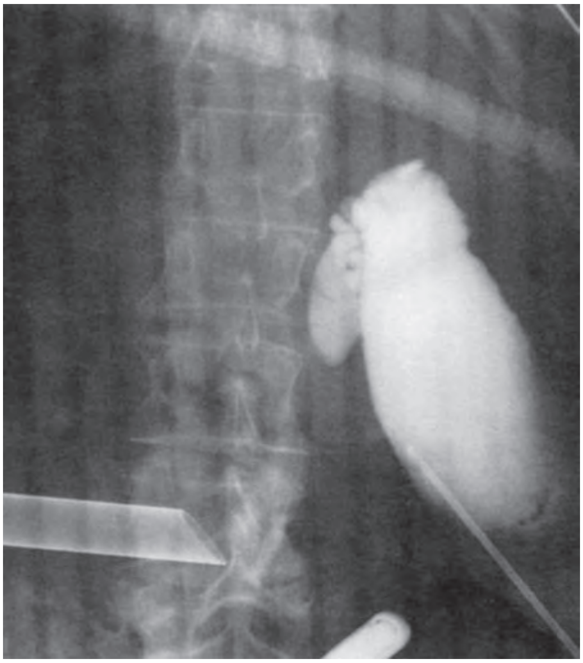


Figure 25.29 Laparoscopic cholecystocholangiogram in a patient with inoperable carcinoma of the pancreas. As the distance between the entry of the cystic duct into the bile duct is more than 1.5 cm from the upper limit of the tumour occlusion, a cholecystojejunostomy will provide adequate palliation of the jaundice and itching during the patient's lifetime.

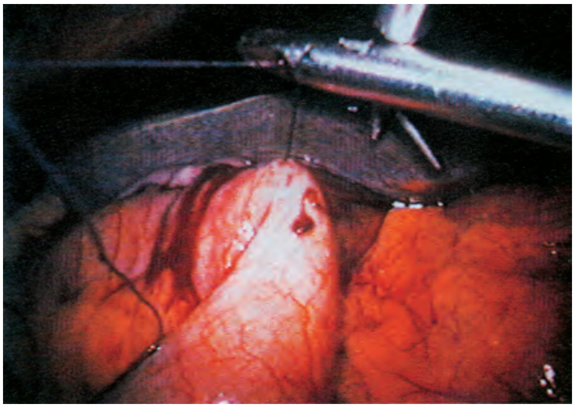


Figure 25.30 Laparoscopic palliative sutured cholecystojejunostomy.

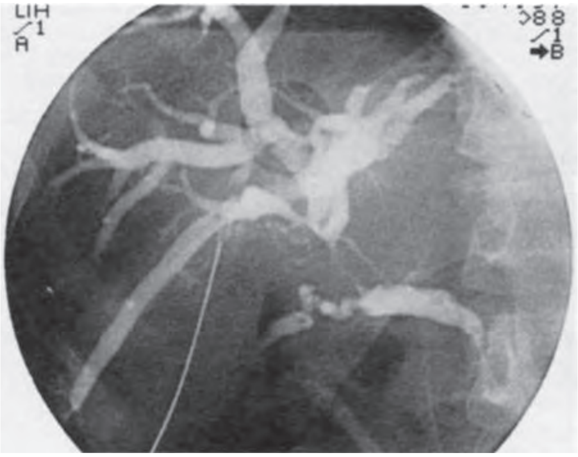


Figure 25.31 Laparoscopic transhepatic cholangiogram showing obstruction of the extrahepatic biliary tract by a large hilar tumour that appeared to originate from the gallbladder.

for jaundice and itching during the patient's lifetime (Figure 25.30). The method involves simple puncture of the gallbladder, using a sharp Veress needle attached by tubing to two 50 mL syringes (saline and diluted contrast). Usually some 40 mL has to be injected before contrast starts exiting the gallbladder into the biliary ducts. At the end of the procedure as much contrast is aspirated as is possible and the puncture site in the gallbladder is sutured with 3/0 absorbable material. However, a choledochoduodenostomy or choledochojejunostomy is preferable as either provide more effective palliative biliary drainage.

Direct puncture

In postcholecystectomy patients, IOFC may be carried out by direct puncture of the common bile duct using a fine intravenous cannula.

Transhepatic operative cholangiography

This is a valuable method that provides very useful information in patients with hilar tumours and during surgery for iatrogenic high bile duct injuries. A Chiba needle is used to puncture an intrahepatic bile duct under contact ultrasound control. When bile is aspirated, contrast is injected into the intrahepatic biliary tree. The technique can also be undertaken laparoscopically (Figure 25.31).

Choledochoscopy (cholangioscopy)

Operative choledochoscopy is well established in biliary tract surgery and is considered an integral part of common bile duct exploration. Two types of choledochoscopes are available: the flexible fibreoptic instrument and the rigid Berci–Shore choledochoscope, which incorporates the Hopkin's rod lens system (Figure 25.32). This is however largely of historical interest as it is more difficult to use and requires a choledochotomy. Furthermore, the rigid choledochoscope can only be used in open surgery. For these reasons it has been replaced by flexible choledochoscopy.

Flexible choledochoscopes

Flexible choledochoscopes do not require preliminary Kocherization of the duodenum for full inspection of the biliary tract and can be passed proximally to visualize the secondary and tertiary intrahepatic ducts. More importantly, they can be used in both open and laparoscopic biliary surgery. The standard choledochoscope has an external diameter of 4.0–4.5 mm and was designed for insertion into the common bile duct through a small choledochotomy (open or laparoscopic). The smaller ones (outer diameter of 2.8–3.0 mm) are referred to as mini-choledochoscopes and are used for transcystic duct extraction of ductal calculi during laparoscopic surgery. Often this entails prior dilatation of the cystic duct with a cylindrical balloon dilator. All mini-choledochoscopes used for laparoscopic extraction of ductal calculi through the cystic duct approach must have an instrument channel of 1.0 mm diameter or larger

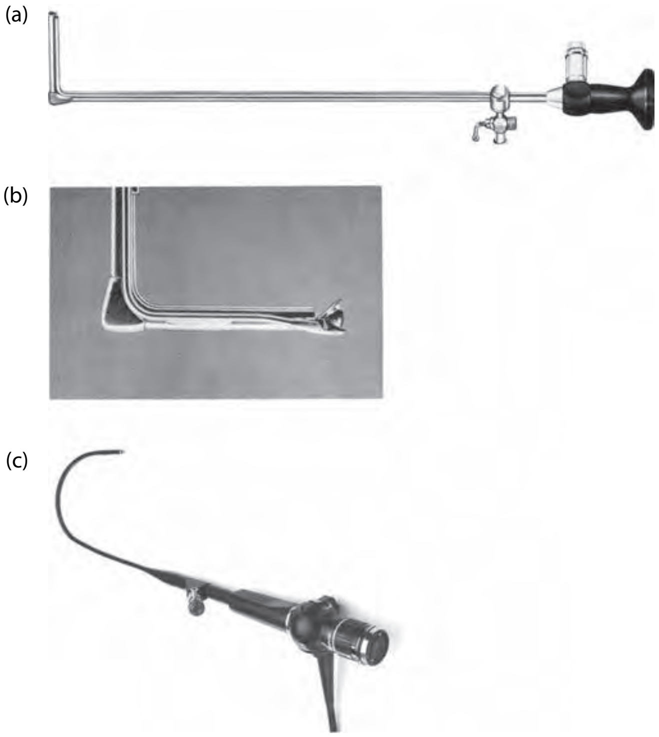


Figure 25.32 (a) The rigid choledochoscope (Berci–Shore). The main disadvantage of this rigid choledochoscope is that it requires a choledochotomy for its insertion and can only be used in open surgery. It has an attachable instrument channel through which biopsy forceps, biliary balloon catheters and Dorian baskets can be passed into the bile duct for visually guided procedures such as stone extraction and biopsy of lesions. (b) It has an attachable strong stone forceps which moves together with the instrument and is extremely useful for dealing with impacted stones at the lower end of the bile duct. The stone can be dislodged and retrieved or crushed under vision. (From Cuschieri and Berci, *Common Bile Duct Exploration*. Dordrecht, The Netherlands: Martinus Nijhoff, 1984, with permission.) (c) The Olympus flexible choledochoscope.

to enable insertion of small wire baskets (3 Fr) and balloon catheters and at the same time leave enough space for effective irrigation. During laparoscopic choledochoscopy, the endoscope is attached to a second charge-coupled device (CCD) camera and with a picture-in-picture display (if available) both the laparoscopic and the endoscopic image are seen on the one monitor. This dual imaging greatly facilitates manipulations, especially extraction of ductal calculi.

The advantages of choledochoscopy are:

- it provides the best evaluation of intracholedochal pathology
- it allows biopsy of suspicious lesions (Figure 25.33)
- it enables visually guided extraction of floating and impacted ductal calculi during both open and laparoscopic surgery
- the routine use of completion choledochoscopy results in an almost negligible incidence of retained ductal calculi
- it reduces the incidence of trauma, especially to the lower end of the bile duct caused by blind instrumentation with metal sounds and forceps
- the flexible endoscope provides an effective method of stone extraction of retained ductal calculi through the T-tube tract.

Biliary manometry

Biliary pressure studies can be undertaken intraoperatively or endoscopically during ERCP using a special perfusion catheter.

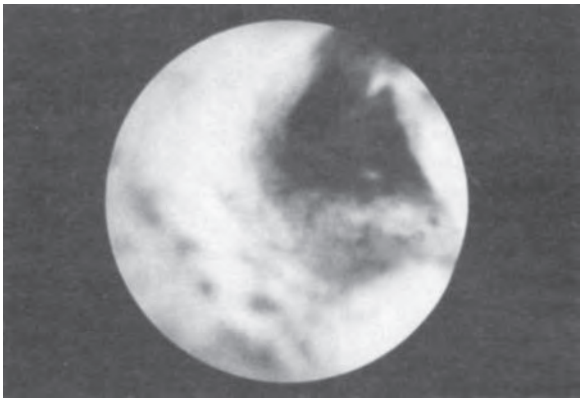


Figure 25.33 Choledochoscopic view of a stenosing lesion of the common hepatic duct. An endoscopic biopsy confirmed the presence of a cholangiocarcinoma.

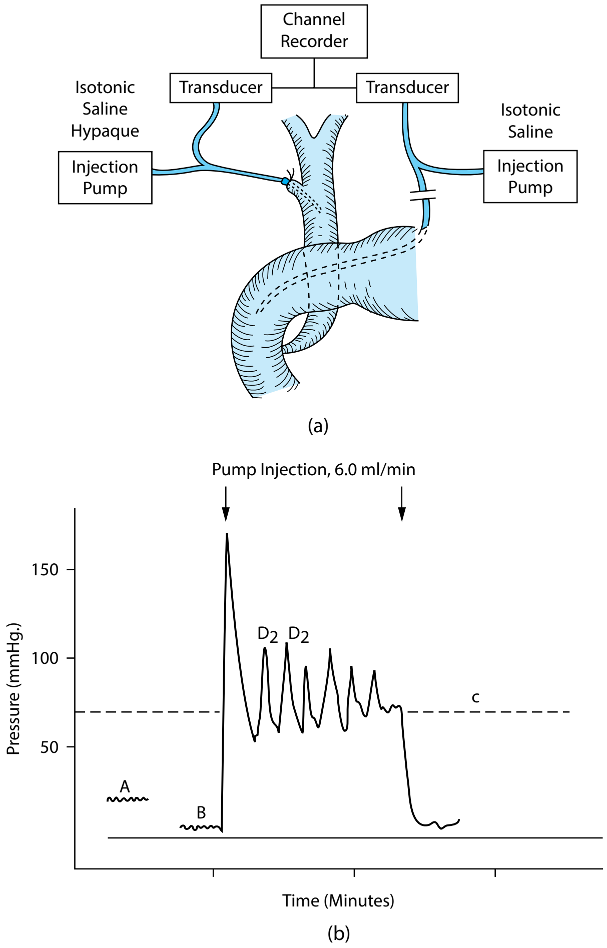
Intraoperative biliary manometry

This can be performed by the use of a pressure transducer connected to the cannula, which is inserted into the common bile duct via the cystic duct. The transducer is attached to a channel recorder which gives an instant display of the biliary pressure. This technique is known as radiomanometry and permits the measurement of the basal (resting) pressure and the filling pressure during the constant infusion of saline (5.0 mL/min). In addition, it demonstrates the sphincteric contractions and may be used to document the effect of spasmolytic drugs on the choledochal sphincter in patients with biliary dyskinesia (Figure 25.34). Operative biliary manometry is, however, seldom used nowadays except for research purposes.

The other technique used during biliary tract surgery is known as manodebimetry. This measures the passage (yield) pressure at the choledochal sphincter and the flow rate through the common channel into the duodenum. The technique was first described by Caroli (Figure 25.35). The modern modification of the Caroli instrument is known as the Tondelli manodebitometer. After measurement of the passage pressure, the upper limit of which is 25 cm H_2O, the calibrated reservoir (filled with saline or contrast medium) is raised to a standard height of 30 cm above the level of the common bile duct and the flow rate through the common channel measured from the rate of emptying of the reservoir per unit time. The normal flow rate measured in this way should exceed 12.0 mL/mm. A high passage pressure and a diminished flow rate are indicative of obstructive disease.

Endoscopic biliary manometry

Endoscopic biliary manometry is now an established diagnostic procedure in specialized units. It is performed by the use of a special perfusion catheter attached to an external transducer and has been used in the investigation of patients with persistent symptoms and pain after cholecystectomy in an attempt to characterize abnormalities of the sphincter (stenosis, dyskinesia). The procedure measures the basal sphincter pressure, the rate and propagation of sphincteric contractions and the response to pharmacological agents such as morphine and CCK. In patients with dyskinesia, increased basal pressure, altered frequency and amplitude of phasic contractions, and reversal of the normal peristaltic direction (retrograde propulsion) have been reported.



(a)

(b)

Figure 25.34 Peroperative radiomanometry: (a) apparatus; (b) normal pressure profile of the *common* bile duct during constant pump infusion of isotonic saline. A, pressure generated by the pump and intrinsic resistance of the system; B, basal bile duct pressure; C, filling pressure during injection; D, contractions of the sphincter of Oddi.

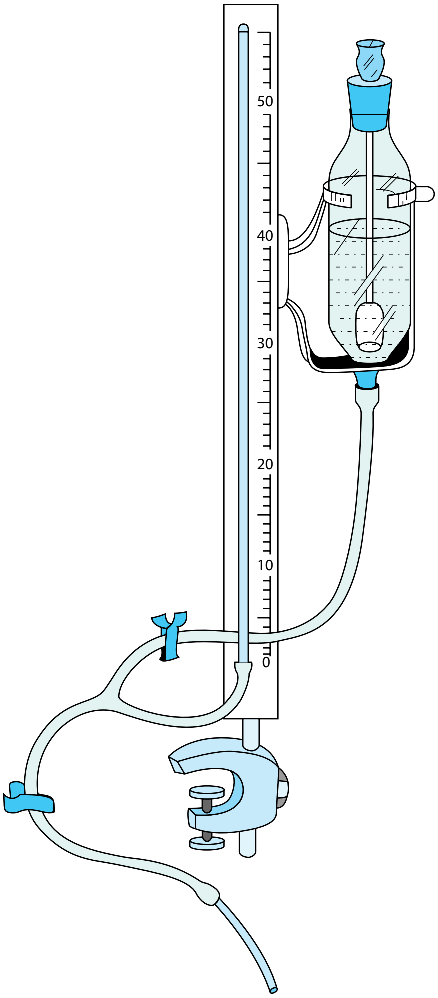


Figure 25.35 The Caroli instrument for manodebimetry. The technique measures the passage (yield) pressure of the choledochal sphincter and the flow rate through the common channel into the duodenum.

Laparoscopy

This procedure should be used routinely by surgeons in all patients with jaundice of malignant origin. It usually provides a direct visualization of the underlying pathology and the exact cause of the jaundice can be determined in all patients if ancillary techniques such as laparoscopic ultrasonography, cholangiography (transhepatic or transcholecystic) and targeted biopsy or cytology are employed. The liver, gallbladder, extrahepatic biliary tract and pancreas are directly visualized, as is the peritoneal lining and most of the intraperitoneal contents. Aside from detecting hepatic disease, primary neoplasms, secondary tumour deposits in the liver and peritoneal dissemination, all of which can be biopsied for histological confirmation, laparoscopy is invaluable in the staging of hepatobiliary and pancreatic tumours. Thus it avoids unnecessary laparotomy in patients with inoperable disease and can now be used to construct laparoscopic bilioenteric and gastrojejunal bypass procedures for advanced inoperable pancreatic cancer. In patients with chronic liver disease and a bleeding tendency, laparoscopic liver biopsy is undertaken in preference to the blind percutaneous procedure as bleeding from the biopsy site can be controlled by electrocoagulation.

Jaundice

Bilirubin is produced in the reticuloendothelial system from the enzymic breakdown of haem, which is derived from effete red blood corpuscles. As it is water insoluble, bilirubin is carried bound to albumin in the plasma and is taken up by the hepatocytes by means of specific membrane carriers. Within the hepatocytes, the bilirubin is stored bound to specific binding proteins (ligandins Y, Z) and then conjugated by a specific enzyme (glucuronyl transferase) to the water-soluble bilirubin glucuronide (conjugated bilirubin) that is then secreted by means of specific carriers into the bile canaliculi, and finally excreted into the biliary tract and intestine. Bacterial degradation of some of the excreted conjugated bilirubin in the distal small bowel results in the formation of urobilinogen, which is reabsorbed and subsequently excreted in the urine and bile. The normal upper limit of serum bilirubin is 17 µmol/L.

Jaundice (hyperbilirubinaemia) is a syndrome of varied aetiology which may be recognized clinically when the serum bilirubin exceeds 40 µmol/L. The hyperbilirubinaemia may be either conjugated or unconjugated and may result from:

- excess bilirubin production
- impaired uptake by the hepatocyte
- failure of conjugation
- impaired secretion of conjugated bilirubin into the bile canaliculi
- impairment of bile flow subsequent to the secretion by the hepatocytes – *cholestatic or obstructive jaundice.*

The defect may be congenital (benign congenital hyperbilirubinaemias) but much more commonly it is acquired as a result of haemolysis, liver disease, adverse drug reaction and biliary tract obstruction which may be intra- or extrahepatic. The early diagnosis and prompt treatment of patients with jaundice reduces both the morbidity and mortality of the underlying disease. In clinical practice, the largest groups by far are those with hepatocellular and cholestatic jaundice.

Hepatocellular jaundice

This is due to parenchymatous liver disease which may be acute (viral hepatitis, liver cell necrosis, acute alcoholic hepatitis, etc.) or chronic (chronic active hepatitis, the various types of cirrhosis, primary biliary, etc.). The principal defect is the failure of secretion of the conjugated bilirubin into the bile canaliculi. The serum transaminases are grossly elevated especially in acute disease. In patients with alcohol-related liver disease, the γ-glutamyl transpeptidase (γ-GT) is elevated as a result of microsomal induction rather than cholestasis. Acute hepatitis due to viral infection or drugs may also cause a cholestatic picture, in which case the alkaline phosphatase and 5-nucleotidase are elevated. The hyperbilirubinaemia is always (predominantly) of the conjugated variety with the presence of bilirubin in the urine even in the absence of a cholestatic component.

Cholestatic jaundice

This is the result of impaired bile flow to the duodenum subsequent to the secretion of conjugated bilirubin into the bile canaliculi. The block may be intrahepatic when it may be functional (e.g. drugs, hepatitis) or organic (obstruction of the intrahepatic biliary tree) or extrahepatic, also known as large bile duct obstruction which constitutes the most important surgical subgroup of cholestatic jaundice as it is always the result of organic disease, e.g. ductal calculi, pancreaticobiliary cancer, etc.

The biochemical features of cholestasis are:

- conjugated hyperbilirubinaemia
- elevation of alkaline phosphatase, 5'-nucleotidase and γ-GT; the enzyme 5'-nucleotidase is the most reliable since its level is not influenced by bone disease and the enzyme is not induced by alcohol
- minimal or no elevation of the serum transaminases
- presence of bilirubin in the urine as the conjugated bilirubin is water soluble and is therefore filtered in the glomerulus
- elevation in the serum cholesterol and bile acid levels although these are not routinely measured in patients with cholestatic jaundice.

It is important to stress that the above biochemical markers of cholestasis do not distinguish between intra- and extrahepatic obstruction.

Haemolytic jaundice

The unconjugated hyperbilirubinaemia results from excess haemolysis. Bilirubin is not present in the urine as the unconjugated pigment is water insoluble and is carried in the plasma bound to albumin. The excess bilirubin production is accompanied by an increased secretion of the conjugated pigment in the bile and therefore increased production of urobilinogen by bacterial decomposition in the distal small intestine. The urine, therefore, contains an excess amount of urobilinogen and urobilin. A cholestatic component may develop in patients with prolonged and recurrent haemolysis (e.g. congenital haemolytic anaemias).

In some patients excess bilirubin production is present in the absence of overt haemolysis. The excess unconjugated bilirubin is thought to result from breakdown of precursor/immature red cells in the bone marrow. This form of benign non-familial congenital hyperbilirubinaemia is referred to as *shunt hyperbilirubinaemia.*

Benign familial congenital hyperbilirubinaemias

This group includes Gilbert disease, Dubin–Johnson syndrome and the Rotor syndrome. All three conditions are congenital and familial. Gilbert disease is due to a defect in the uptake of bilirubin by the hepatocytes and results in mild unconjugated hyperbilirubinaemia. Both the Dubin–Johnson and Rotor syndromes are caused by a secretory defect of conjugated bilirubin by the hepatocytes into the bile canaliculi and therefore lead to a conjugated hyperbilirubinaemia. In addition, patients with the Dubin–Johnson syndrome are unable to excrete contrast medium into the biliary tree and, for this reason, the gallbladder is not visualized by oral cholecystography and intravenous cholangiography. Despite the accumulation of conjugated bilirubin in the blood and its appearance in the urine, there are no other biochemical markers of cholestasis in both conditions.

Management of patients with large bile duct obstruction

It is important to reiterate that a properly taken history and physical examination will allow a correct diagnosis to be made in some 80% of patients. Surgical obstructive jaundice (or large bile duct obstruction) is always accompanied by dilatation of the biliary tract. In essence, the management entails establishing the cause, the general condition of the patient and, in the case of tumours, the stage of the disease. Malignant large bile duct obstruction may be inoperable either because the lesion is not resectable or because the patient's ASA grade or POSSUM risk assessment score precludes major surgical intervention. In this situation management is directed towards palliation by endoscopic/radiological stenting or laparoscopic bypass procedures.

Dilatation of the biliary tract

The common bile duct dilates more rapidly than the intrahepatic biliary tree. Thus, instant passive dilatation of the common bile duct demonstrated fluoroscopically can be produced by excessive filling of the biliary tract by contrast medium. Dilatation of the intrahepatic biliary tree always signifies prolonged obstruction and, experimentally, it requires a minimum of 3 weeks of obstruction for the production of demonstrable intrahepatic duct dilatation. By convention, during surgery (as opposed to ultrasound examination; see section External abdominal ultrasonography) the diameter of the common bile duct is measured just above the junction of the cystic duct. A common bile duct whose diameter exceeds 10 mm after contrast injection (during cholangiography) is considered dilated. In the absence of contrast injection, e.g. ultrasound or MRCP, a diameter above 8 mm indicates dilatation.

Dilatation of the common bile duct signifies existing or recently relieved obstruction, the most common cause of which is calculous disease. There is an established positive correlation between the duct diameter and the incidence of ductal stones. On the other hand, stones may be present in a normal-sized common bile duct and in several reported series a 5–10% incidence of ductal stones has been reported in patients with common bile ducts of 5 mm. Other causes of duct dilatation include pancreaticobiliary cancer, chronic pancreatitis, congenital cystic disease and parasitic infestation. Controversy persists regarding dilatation of the bile duct after truncal vagotomy. Some studies have demonstrated abnormal gallbladder emptying after this procedure but not following highly selective vagotomy. The same controversy surrounds the view that the diameter of the common bile duct increases with age.

Although there are no intrahepatic communications between the right and the left intrahepatic ductal trees in the normal state, communications between the two systems develop following the onset of extrahepatic large bile duct obstruction.

Preoperative management of the patient with obstructive jaundice

Adequate timing of the surgical intervention and preparation of the patient for surgery are essential in the management of patients with obstructive lesions of the biliary tract. Undue delays exceeding 3–4 weeks increase both the morbidity and mortality rates following surgical intervention. Adequate preparation entails the correction of metabolic abnormalities, improvement of the general condition and the institution of specific measures designed to minimize the incidence of complications associated with prolonged or severe cholestasis. These include:

- infections: cholangitis, septicaemia, wound infections
- disorders of the clotting mechanism
- renal failure
- liver failure
- fluid and electrolyte abnormalities.

Furthermore, the conjugation and metabolism of drugs and anaesthetic agents is impaired because of the hepatocyte malfunction. Contrary to popular belief, there is no evidence to support the view that wound healing is impaired in the presence of jaundice. Wound-healing problems are largely confined to patients with malignant obstruction and are the result of the underlying disease and its association with a poor nutritional state. The nutritional deficits of these patients vary with the underlying pathology, age and social class of the patient. In general, parenteral nutrition should be used selectively, and only in those patients who are grossly malnourished, because of infective risks. A high intake of carbohydrates is essential and amino acid solutions containing aromatic amino acids (phenylalanine, tyrosine and tryptophan) should be used sparingly as these may precipitate encephalopathy in susceptible patients. An oral diet is the safest and should be used whenever possible. It may be supplemented by elemental diets in nutritionally compromised patients. Enteral nutrition is considered essential for the maintenance of the mucosal integrity of the gut against bacterial translocation. Some advocate the oral administration of bile salts or lactulose to reduce the intestinal absorption of endotoxin from the intestinal microflora and thus minimize the incidence of renal failure following surgical intervention.

Hypokalaemia is frequently present and should be corrected. In general, intravenous isotonic saline administration should be restricted in the jaundiced patients and those with liver disease as the total exchangeable body sodium is elevated. The low normal values of the concentration of the serum sodium frequently encountered in these patients are due to expansion of the intra- and extravascular fluid compartments consequent on the excessive retention of water (dilutional hyponatraemia). A viral screen is necessary in these patients and, when the serology is positive, special precautions must be taken both in the ward and in the operating theatre to avoid spread of the infection to the attending medical staff.

Prevention of infective complications

Whereas the normal biliary tract and bile in humans is sterile, bacteria are frequently present in biliary tract disorders and may lead to septic complications, particularly cholangitis and septicaemia. Infection of the biliary tract is much more commonly present in ductal calculous disease than in patients with malignant obstructive jaundice. Anaerobes are less frequently found in the biliary tract and duodenum than aerobic bacteria even in the presence of pathological states. Thus, in the absence of stenting, the majority of infections associated with biliary tract disorders are aerobic in origin and most commonly due to Gram-negative bacilli. Endoscopic stenting of patients with malignant large bile duct obstruction results in infection of the biliary tract and is unwise if the patient is deemed operable.

A number of clinical trials have shown that the postoperative sepsis in patients having biliary tract surgery is generally due to bacteria in the bile and the use of short-term prophylactic antibiotics (three-dose regimen perioperatively: immediately before surgery to 24 hours later) significantly lowers the incidence of sepsis only in patients who have bacteria in the bile at the time of surgery. Prophylactic antibiotics are therefore not advised in all patients undergoing surgery on the biliary tract but should be administered to those patients who are likely to have bacteria in the bile. The higher risk groups have been identified and include:

- all jaundiced patients
- patients with rigors and pyrexia
- patients undergoing emergency biliary procedures/operations
- elderly patients
- patients with common duct stones even if not jaundiced
- patients undergoing secondary biliary intervention.

Some advocate the use of Gram's staining of the bile at the time of surgery to determine both the presence and the Gram-staining characteristics of the organisms in deciding on the appropriate antibiotic. The use of prophylactic antibiotic therapy with a cephalosporin, or aminoglycoside or pipericillin (three doses) in the high-risk groups outlined above has been shown to reduce the incidence of postoperative wound infection, cholangitis and septicaemia.

Bacterial proliferation in the bile following exploration of the common bile duct and insertion of a T-tube is extremely common and may become a source of infection or lead to the formation of calcium bilirubinate stones as a result of the deconjugation of the bilirubin glucuronide by glucuronidase-producing bacteria, particularly *Escherichia coli*. Thus, a closed system of T-tube drainage should always be used and a bile culture performed a few days before the removal of the T-tube. A course of the appropriate antibiotic should be administered if the culture is positive.

Correction of disorders of coagulation

The most common disorder of coagulation encountered in patients with large bile duct obstruction is a prolonged prothrombin time resulting from a deficiency of vitamin K-dependent factors consequent on the malabsorption of this vitamin which occurs in cholestatic jaundice. The intramuscular injection of phytomenadione (10–20 mg) will reverse the multifactorial clotting deficiency within 1–3 days. Severe hepatic disease, usually with a poor prognosis, is present if the prothrombin time remains abnormally prolonged despite this treatment. If these patients require surgical intervention, administration of fresh frozen plasma is necessary to cover the perioperative period.

A more serious bleeding disorder may arise usually in the severely jaundiced patient who may develop a consumptive coagulopathy from a disseminated intravascular coagulation due to the presence of circulating endotoxin. This serious haematological complication requires careful monitoring of fibrinogen levels, fibrinogen degradation products and platelet counts. It may improve with control of the infection but often requires specific treatment with fresh frozen plasma alone or in combination with heparin.

Prevention of renal failure

The association between postoperative renal failure and severe conjugated hyperbilirubinaemia is well known but the underlying mechanism of the renal impairment is inadequately understood, although a reduced glomerular filtration is usually present. Even in the absence of infection, endotoxinaemia is frequently present in jaundiced patients when it results from absorption of endotoxin produced by the intestinal microflora. There appears to be a relationship between impaired renal function and the presence of circulating endotoxin in jaundiced patients. Irrespective of the exact cause of the renal damage, there is now good evidence that adequate hydration and preoperative induction of a natriuresis/diuresis reduces the incidence of renal failure after surgical intervention in jaundiced patients. It is current routine practice to administer intravenous fluids (5% dextrose saline) for 12–24 hours before surgery. This is followed by an osmotic diuretic (mannitol) or a loop diuretic (frusemide) administered intravenously at the time of induction of anaesthesia. All patients undergoing surgery should be catheterized and the urine output measured hourly. Further administration of diuretics (mannitol or frusemide) is indicated if the urine output falls consistently below 40 mL/h (despite adequate hydration and normovolaemia) during operation and subsequently thereafter.

Preoperative administration of oral chenodeoxycholate commencing a few days before surgery is practised in some centres and one clinical trial has shown a reduction in the incidence of renal failure, although a second trial with the epimer ursodeoxycholate did not report any benefit.

Also, the administration of oral lactulose has been shown to reduce the incidence of renal failure in jaundiced patients undergoing surgical treatment but this requires confirmation.

Prevention of hepatic encephalopathy

Liver failure is usually encountered in patients with prolonged complete large bile duct obstruction or those patients with pre-existing chronic hepatocellular disease, such as cirrhosis, chronic active hepatitis, etc., who undergo surgery. If the jaundice is severe (above 150 μmol/L) or the patient shows signs of impending liver failure, a period of decompression is indicated. This is nowadays achieved by insertion of a plastic endoprosthesis for patients with malignant obstruction. Alternatively an endoscopic sphincterotomy is performed in patients with periampullary cancer. External percutaneous decompression via a transhepatic tube draining into an external collecting system is no longer advocated since it predisposes to infection and leads to a loss of bile acids unless the bile is returned to the gastrointestinal tube via a nasogastric tube. The correction of hypokalaemia, the restricted use of sedatives, hypnotics and potent analgesics, and the prompt treatment of infection cannot be overemphasized. If sedation is required, small doses of promethazine or chlorpromazine can be administered.

Postoperative jaundice

Jaundice occurring for the first time in the postoperative period may be due to a variety of causes:

- benign reactive hepatitis – self-limiting
- sepsis – leaking anastomosis, abscess formation, septicaemia, pneumonia, etc.
- pre-existing primary disease – known or suspected liver disease
- major hepatic resections
- drug-induced liver disease
- biliary pathology – residual stones, bile duct trauma and missed tumours/ pathology of the biliary tract
- massive blood transfusion – each unit of blood provides a bilirubin load transfusion of 250 mg/100 mL when fresh
- haemolysis – unconjugated hyperbilirubinaemia, positive Coombs' test
- residual haematomas – unconjugated hyperbilirubinaemia.

It always requires detailed investigation to establish the cause and outline the necessary course of action. Mild self-limiting conjugated hyperbilirubinaemia, sometimes referred to as benign postoperative cholestasis, may follow prolonged operations and fever caused by chest infections. It is caused by a reactive hepatitis, which is probably multifactorial in origin, resulting from a combination of reduced liver blood flow, hypoxia, hypercarbia, breakdown of transfused cells and temporary hepatocellular dysfunction. Initially, it warrants observation with repeated assessment of the patient's condition and sequential biochemical profiles, as the majority of cases with this syndrome subside within 3–4 weeks.

Marked cholestatic jaundice develops after extensive hepatic resection, especially right hepatectomy and extended right hepatectomy. The serum bilirubin rises over a period of several days to reach a plateau at 7–10 days, after which it should start to decline. In view of the extensive hepatic resection (low residual liver parenchyma), the rise in the serum alkaline phosphatase is small.

Severe or progressive jaundice in the postoperative period is always sinister and usually indicates a primary biliary tract problem, or significant liver disease or severe sepsis such as that resulting from an anastomotic dehiscence. Aside from the usual liver function tests, the following may be required:

- blood culture
- ultrasound/CT scanning for the detection of collections/bilomas/ abscesses
- biliary scintiscanning
- MRCP
- contrast radiological examination of recently constructed gastro-intestinal anastomoses
- liver biopsy if the above are negative or if hepatocellular disease or drug-induced jaundice is suspected.

Hepatitis B is rare nowadays because of better screening for the viral antigens and improved blood donor selection. Most instances of infection caused by blood and blood products are due to hepatitis C virus and other non-A, non-B viruses. Drug-induced jaundice is common in hospital practice. Some of the important drugs which may give rise to this adverse reaction are listed in Table 25.1.

Severe hepatotoxicity can follow halothane anaesthesia. In the majority of patients (over 80%), this follows repeated exposure usually within 28 days (75%). The disease is usually severe and is accompanied by an overall mortality of 40%. The following recommendations have been issued by the Committee on the Safety of Medicines.

- A careful history to determine previous exposure and reactions to halothane should be obtained from every patient.
- Repeated exposure to halothane within a period of at least 3 months should be avoided if at all possible.
- A history of unexplained jaundice or pyrexia following exposure to halothane is an absolute contraindication to further use of halothane in that individual patient.

Clinical management of postoperative jaundice

In the first instance, a full examination of the patient and a careful reappraisal of the preoperative liver function tests are carried out. If liver function was normal prior to operation, the following are performed in a sequential order.

- *Determination of the nature of the jaundice by the liver function tests and urine testing for bilirubin* – postoperative unconjugated hyperbilirubinaemia is commonly due to large/massive transfusion. Each unit of blood provides a bilirubin load of 250 mg/100 mL of fresh blood. Unconjugated hyperbilirubinaemia may also result from resorption of residual haematoma/haemoperitoneum or haemolysis. Haemolytic reactions resulting from minor/major incompatibilities are accompanied by systemic signs. When suspected, screening tests for haemolysis should be performed.

Table 25.1 Drug-induced liver damage

Category	Example	Hepatic lesion
Antibiotics	Tetracyclines, especially i.v. – dose related Penicillins – hypersensitivity Chloramphenicol Sulphonamides – hypersensitivity	Fatty infiltration Hepatitis Hepatitis Granulomas, focal hepatocellular necrosis
Analgesics and anti-inflammatory drugs	Paracetamol – dose dependent Phenylbutazone – hypersensitivity Carbamazepine Salicylates – dose related	Centrilobular necrosis, massive liver necrosis Hepatitis with granuloma may progress to cirrhosis Cholestasis Focal hepatic necrosis
Psychotropic drugs	Monoamine oxidase inhibitors Phenothiazines – hypersensitivity Tricyclic antidepressants	Hepatitis, may progress to massive hepatic necrosis Hepatitis and cholestasis Cholestatic hepatitis, more usually mild elevation of transaminases
Steroids	Testosterone and anabolic steroids Oestrogens	Cholestasis, peliosis hepatis, hepatic tumours Cholestasis, gallstones, hepatic tumours
Anaesthetic agents	Halothane	Hepatitis which may progress to massive liver necrosis
Antituberculous drugs	PAS – hypersensitivity INAH – related to acetylator status Rifampicin – dose related	Hepatitis Focal to severe hepatic necrosis Defective bilirubin transport, mild hepatitis
Cytotoxic and immunosuppressive drugs	Azathioprine 6-Mercaptopurine – dose related Methotrexate – long-term therapy	Cholestasis and peliosis hepatis Hepatitis Fatty change, fibrosis of the portal tracts and cirrhosis
Others	Benzothiazine diuretics Phenindione Chlorpropamide Phenytoin	Cholestatic hepatitis Cholestatic hepatitis Cholestatic hepatitis Hepatocellular necrosis

i.v., intravenous; INAH; isoniazid; PAS, 4-aminosalicylic acid.

- *If the jaundice is cholestatic*, an assessment of all the drugs and anaesthetic agents used is followed by the withdrawal of any drug known to cause hepatotoxicity. A thorough search for sepsis by the appropriate tests (including blood cultures) is made and an ultrasound/CT examination carried out. External leakage of bile or evidence of anastomotic dehiscence in a jaundiced patient is always serious and establishes the cause. Cholestatic jaundice accompanied by sepsis and abdominal signs is always indicative of a major surgical complication, e.g. bile duct injury, cholangitis or anastomotic dehiscence, and requires urgent radiological investigation. In the absence of sepsis and if the MRCP confirms integrity of the biliary tract, the liver function tests are repeated daily to determine the course of the biochemical profile. If the cholestasis is seen to be resolving, an expectant policy is adopted, otherwise a liver biopsy is undertaken.

Jaundice in infancy and childhood

Apart from the physiological jaundice, the aetiology of hyperbilirubinaemia in infancy may be due to haematological disorders, enzymatic defects, inborn errors of metabolism, infections and obstructive disease. The causes of jaundice in infancy and childhood are:

- physiological jaundice
- haematological disorders – inspissated bile plug
- enzymatic defects
- inborn errors of metabolism
- hepatitis and other infections
- BA
- biliary hypoplasia
- cystic disease of the bile ducts
- congenital perforation of the common bile duct.

Whereas unconjugated hyperbilirubinaemia may be physiological, all causes of conjugated hyperbilirubinaemia are abnormal. Inflammatory disease of the gallbladder and cholelithiasis are rare but do occur, including acalculous cholecystitis. A higher percentage of gallstones in the paediatric age group are associated with haemolytic disorders than in the adult population. An increased incidence of gallstones is also found in patients with cystic fibrosis. Children with gallstones present with an atypical history of vague abdominal pain and distress. Classical biliary colic is rare. An ultrasound examination of the gallbladder should be performed in all children undergoing splenectomy for haemolytic anaemia.

Biliary atresia

BA (also known as progressive obliterative cholangiopathy) is a disease of neonates/infants which is characterized by obliteration of the extrahepatic biliary system, resulting in large bile duct obstruction. It is the most common surgically treatable cause of cholestasis in the newborn. Untreated, BA leads invariably to secondary biliary cirrhosis within 2–3 months of onset, which indicates the importance of early diagnosis and decompression of the obstructed biliary tract.

Patients with BA fall into two groups: (1) those with isolated BA (postnatal form accounting for 65–90% of cases) and

(2) infants with associated situs inversus or polysplenia/asplenia with or without other congenital anomalies (fetal/embryonic form) comprising the rest (10–35%). In terms of the pathological anatomy three types are distinguished:

- type I in which the common bile duct is obliterated but the proximal ducts are patent
- type II in which the atresia affects the hepatic duct with cystic structures being found in the porta hepatis
- type III (the commonest, 90% of patients) in which there is atresia of the right and left hepatic ducts up to the level of the porta hepatis.

Pathology

The pathogenesis of BA remains unknown. There is no evidence to support the view that BA is a congenital malformation of the biliary ductal system. The current opinion on its aetiology favours an acquired origin although no single factor has been identified. Infectious agents are regarded as the most plausible candidates and some studies have identified elevated antibody titres to reovirus type 3 in patients with BA when compared with controls. Other viruses have been implicated including rotavirus and cytomegalovirus. Although BA was initially considered as a cholangiopathy induced by disordered bile acid metabolism, there is no evidence for this, but retained bile acids are thought to be important in the progressive damage to the intrahepatic ducts.

In type III atresia (the most prevalent), a fibrous remnant indicates complete obliteration of a segment of the extrahepatic biliary system. The intrahepatic ducts extending to the porta hepatis are initially patent during the first few weeks, but are progressively destroyed presumably by the same agent acting together with retained toxins in bile and other contributing factors. The destruction is thought to result from the release by the biliary epithelial cells of various cytokines, e.g. CCL-2, MPC-1, tumour necrosis factor, interleukin-6, transforming growth factor-β, endothelin and nitric oxide. This cytokine cascade is held to be the ultimate cause of the progressive fibrosis leading to secondary biliary cirrhosis with the development of portal hypertension, oesophageal varices (which may bleed) and hypersplenism.

Clinical features

Although BA occurs in the West with an overall incidence in the USA of 1 per 10 000–15 000 live births, the incidence of the disease is highest in Asian populations, especially Chinese and to a lesser extent Japanese infants. Additionally the disorder is twice as common in black than white infants. BA is also commoner in female infants. The fetal/perinatal form is evident within the first 2 weeks of life, whereas the postnatal type presents in infants aged 2–8 weeks.

A high index of suspicion is key to making an early diagnosis which is essential for the prevention of irreversible biliary cirrhosis. The diagnostic protocol carried out in these infants is designed to differentiate, in the first instance, neonatal hepatitis from BA and the standard work-up, apart from liver function tests, includes screening for infectious, genetic and enzymatic disorders. Studies required include the biliary scintiscanning (no bile passage into the duodenum over 3–4 days in BA as assessed by faecal radioactivity), liver biopsy and serial total and conjugated bilirubin levels. A rising or flat bilirubin curve over several weeks is found in atresia, whereas in neonatal hepatitis there is a gradual fall in the serum bilirubin after an initial peak.

The clinical features include variable degrees of jaundice, dark urine and pale stools. Most affected infants are full term. This is important as it is in contrast to infants with idiopathic neonatal hepatitis, who are often premature or small for gestational age or both, although a higher incidence of low birthweight has been reported in some BA neonates. Appetite, growth and weight gain may be initially normal during the first few weeks of life. Hepatomegaly may be present early, and the liver is firm on palpation. Splenomegaly is common, and an enlarging spleen suggests progressive cirrhosis with portal hypertension. In infants with the fetal/neonatal form, the liver is palpated in the epigastrium and cardiac murmurs indicative of associated cardiac anomalies may be present. BA is characterized by conjugated hyperbilirubinaemia (defined as any level exceeding either 2 mg/dL or 20% of total bilirubin). Infants with BA typically show only moderate elevations in total bilirubin (commonly 6–12 mg/dL), with the conjugated fraction accounting for 50–60% of total serum bilirubin. Alkaline phosphatase, 5'-nucleotidase, γ-GT and serum bile acids are elevated. Aminotransferase levels are not particularly helpful in establishing a diagnosis, although a markedly elevated alanine aminotransferase level (>800 IU/L) indicates significant hepatocellular injury and is more consistent with the neonatal hepatitis.

Ultrasonography is used to exclude specific anomalies of the extrahepatic biliary system, e.g. choledochal cysts. Although, in BA, ultrasonography may demonstrate absence of the gallbladder and no dilatation of the biliary tree, its sensitivity and specificity for the diagnosis of BA is low and thus this imaging is not very useful. In contrast, hepatobiliary scintiscanning is very useful in evaluating infants with suspected BA as it provides unequivocal evidence of either patency with intestinal excretion of radiolabel (excluding BA) or obstruction of the extrahepatic biliary system, confirming the diagnosis. However, the reliability of the scintiscan is diminished in patients with very high conjugated bilirubin levels (>20 mg/dL). Additionally, the test has a 10% false-positive or false-negative rate. Percutaneous liver biopsy provides valuable information. When examined by an experienced specialist pathologist, it can differentiate between obstructive and hepatocellular causes of cholestasis, with 90% sensitivity and specificity for BA.

The important differential diagnosis is between BA and neonatal hepatitis and *Alagille syndrome*. The latter is an autosomal dominant disorder (OMIM 118450) associated with abnormalities of the liver, heart, skeleton, eye and kidneys and a characteristic facial appearance. The abnormality of the liver consists of hypoplasia of the hepatic ducts which is not surgically correctable.

Treatment

The treatment of BA is surgical and must be carried out within 60 days of birth, otherwise the prognosis is poor from irreversible cirrhosis. At operation a cholecystocholangiogram (if the gallbladder is present) and a wedge liver biopsy are performed.

If an extrahepatic stump of the common hepatic duct is present, a Roux-en-Y jejunal anastomosis is carried out. Hepatic portoenterostomy (the Kasai procedure) is performed if no extrahepatic ducts are discernible. The procedure consists of the progressive excision of fibrosed remnants of the ducts anterior to the portal vein at the porta hepatis together with a 1 cm ring of adjacent liver substance, advancing to some 2–3 cm in depth using the operating microscope until biliary structures are identified: bile ducts, collecting tubules of biliary glands or biliary glands. The excised scar tissue is subjected to histological examination to identify these structures. A modified Roux-en-Y anastomosis (Suruga II procedure) is then performed at the periphery of the saucerized area surrounding the bile ductules (Figure 25.36). This has an access jejunostomy placed subcutaneously which allows irrigation, evaluation of postoperative bile flow and, if necessary, introduction of a paediatric flexible endoscope to inspect the porta hepatis.

In an effort at reducing postoperative cholangitis due to reflux of intestinal contents, various valve constructions have been advocated between the portoenterostomy and the enteroenteric anastomosis. Some favour an isoperistaltic jejunal loop with a nipple valve interposed between the porta hepatis and the duodenum instead of the Roux-en-Y reconstruction. An 80% successful outcome is obtained in infants in whom bile ducts communicating with the intrahepatic system have been identified at operation provided this is performed within 60 days of birth. These infants require vitamin E supplements to minimize the development of neurological sequelae. Overall, some 25% of these children develop cirrhosis and portal hypertension. Hepatic transplantation is now accepted as the treatment of choice in infants and children with BA in whom a portoenterostomy has failed and who develop cirrhosis. Recently the argument has been made for proceeding to primary liver transplantation instead of a portoenterostomy on the grounds that the majority of infants treated by the latter procedure develop biliary cirrhosis and portal hypertension with bleeding particularly when a cutaneous stoma leading from the portoenterostomy loop is employed. While primary hepatic transplantation may be appropriate in a small subset with adverse scoring system based on multiple liver function tests, for the majority of infants this policy is generally considered to be inappropriate. The Kasai operation works best at an age less than 6–8 weeks (when the results of hepatic transplantation are not optimal) and imparts long-term control in 30% of patients. Transplantation success improves considerably when the operation is performed after the age of 12 months. Thus a portoenterostomy buys time, allows patient growth, enhances the donor pool available to the patient and increases the chance of successful transplantation. In a large reported experience based on this policy, the 1 year transplant survival rate was 87%. Poor results were encountered only in patients in whom a portoenterostomy had been revised prior to transplantation. Thus a non-functioning portoenterostomy is an indication for hepatic transplantation and attempts at revision are unwarranted. The problem with transplantation in children relates to the scarcity of appropriate-size livers (which can be housed in the small subdiaphragmatic space), since the vast majority of donors are adults. The size disparity has however been largely resolved by recourse to split liver or segmental hepatic transplantation from either brain-dead donors or living related donors. Corticosteroid treatment after a Kasai operation, with or without choleretics and antibiotics, has a beneficial effect on the postoperative bile flow and can clear the jaundice; but there is some debate on dosing and duration of the ideal steroid therapy (high-dose vs low-dose therapy). Prior to hepatic transplantation the only surgical therapeutic option was portoenterostomy when the long-term survival rate for infants with BA was 47–60% at 5 years and 25–35% at 10 years with inevitable death from end-stage biliary cirrhosis and its complications. Hepatic transplantation is the only option for long-term survival in most patients.

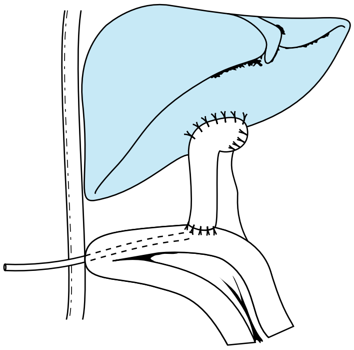


Figure 25.36 The Suruga II procedure of intubated access jejunostomy in patients with biliary atresia.

Biliary hypoplasia

Biliary hypoplasia is not a distinct entity but includes various neonatal disorders which simulate atresia but in whom the expected course of progressive jaundice with the development of biliary cirrhosis does not occur. Indeed in some infants, the jaundice subsides. The conditions include the inspissated bile plug syndrome, partial atresia and infants with small intra- and extrahepatic ducts (true biliary hypoplasia). Some are secondary to another primary disorder, e.g. choledochal cyst, neonatal hepatitis and α_1-antitrypsin deficiency. Early surgical treatment, when indicated, is designed to correct the primary abnormality. The inspissated bile plug syndrome is usually secondary to haemolytic disorders. The diagnosis is made by operative cholangiography and treatment, which is curative, consists of irrigation of the extrahepatic ducts.

Spontaneous perforation of the bile duct

The perforation which occurs at the junction of the cystic with the common duct leads to the formation of a pseudocyst. Surgical treatment consists of transperitoneal drainage, which is usually followed by spontaneous closure of the perforation. Spontaneous perforation of the bile duct (SPBD) is a rare condition affecting infants and children. Most perforations occur close to the junction of the cystic duct and in some patients it leads to the formation of a pseudocyst, whereas in others it causes biliary ascites. SPBD is not considered to have a viral origin but its aetiology is otherwise unknown, although several theories have been proposed:

- congenital mural weakness of the common bile duct
- ischaemia
- distal biliary obstruction
- pancreaticobiliary malunion (PBM).

In some infants there is some evidence to suggest that SPBD and choledochal cysts are related entities with a common pathogenesis to PBM, as intrinsic bile duct weakness and distal bile duct obstruction may be implicated in both conditions. Obstruction distal to the perforation may occur in cases with ductal stones and actual stenosis. However, distal obstruction is commonly secondary to the perforation from stasis and usually resolves with adequate drainage.

Clinical features

SPBD presents during the first to the third month of life. The most common presentation is with abdominal distension due to 'ascites' (60%) and/or jaundice (60%). Other symptoms include fever, vomiting, abdominal mass and pale stools. Less commonly a hernia or a hydrocele may be present, presumably secondary to the increased intra-abdominal pressure. The general condition is relatively good at presentation in most infants. Less commonly SPBD presents acutely with poor general condition requiring resuscitation. The serum bilirubin and liver enzymes are either normal or only mildly elevated. A mild leucocytosis is present and an ascitic tap reveals bile-stained fluid with a bilirubin concentration higher than that in the serum.

Treatment

The treatment is surgical and should be covered by broad-spectrum antibiotic therapy which is continued for a few days postoperatively unless the patient develops cholangitis which necessitates a full course of antibiotic based on culture and sensitivity tests. Intraoperative cholangiography is performed in all of the cases if possible. This is necessary to outline the biliary tract, and document any calculi and distal obstruction. Repair of the perforation is unnecessary, may be hazardous and carries a definite risk of postoperative stricture. Exploration of the porta hepatis is avoided in view of the inflammatory state. The standard treatment consists of insertion of a T-tube through the perforation, lavage and suction of peritoneal contents and insertion of a peritoneal drain with or without cholecystectomy. The distal obstruction resolves after T-tube drainage in the majority of cases. A choledochojejunostomy is indicated at a subsequent stage only in cases where the obstruction persists (on repeat T-tube cholangiography) as it indicates the distal stenosis.

Cystic disease of the biliary tract

Cystic disease of the biliary tract is rare and in the USA accounts for 1 in 13 000 hospital admissions. The aetiology remains unknown. Most instances are thought to represent congenital weakness of the common bile duct with distal obstruction caused by an anomalous acute or right-angle junction between the pancreatic duct and the common bile duct resulting in an abnormally long common channel (>0.6 cm). Another theory postulates an unequal proliferation of the duct epithelium. Several types are recognized (Figure 25.37):

- type I diffuse choledochal cystic dilatation, commonest; B type II localized dilatation of the supraduodenal bile duct
- type III supraduodenal diverticulum
- type IV intraduodenal diverticulum or choledochocele
- type V solitary intrahepatic cyst
- type VI multiple intrahepatic cysts (Caroli disease)
- type VII multiple intra- and extrahepatic cysts.

Type VI (Figure 25.38), which is a hereditary disorder (autosomal recessive), is known as Caroli disease and carries a poor prognosis from recurrent cholangitis and the development of intrahepatic stones, liver abscess formation and, eventually, cirrhosis. In some patients, multiple intrahepatic cystic disease is accompanied by congenital hepatic fibrosis. This condition is known as Grumbach disease.

Clinical features

There appears to be a high incidence of cystic disease in the Japanese. There is a strong female predominance worldwide (70% of reported cases in the West) and 25% of the reported

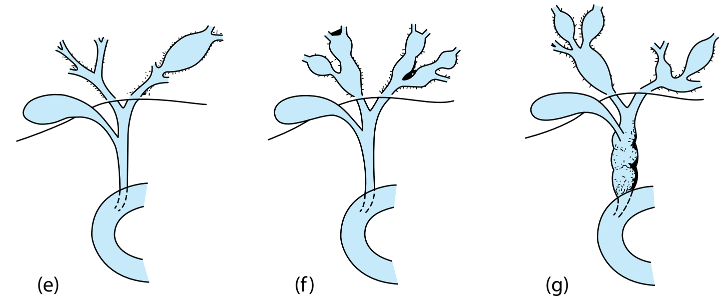


Figure 25.37 Diagrammatic representation of the types of cystic disease of the biliary tract: (a) type I, diffuse choledochal dilatation; (b) type II, localized choledochal dilatation; (c) type III, supraduodenal diverticulum; (d) type IV, intraduodenal diverticulum (choledochocele); (e) type V, solitary intrahepatic cyst; (f) type VI, multiple intrahepatic cysts (Caroli disease); (g) type VII, multiple intra- and extrahepatic cysts.

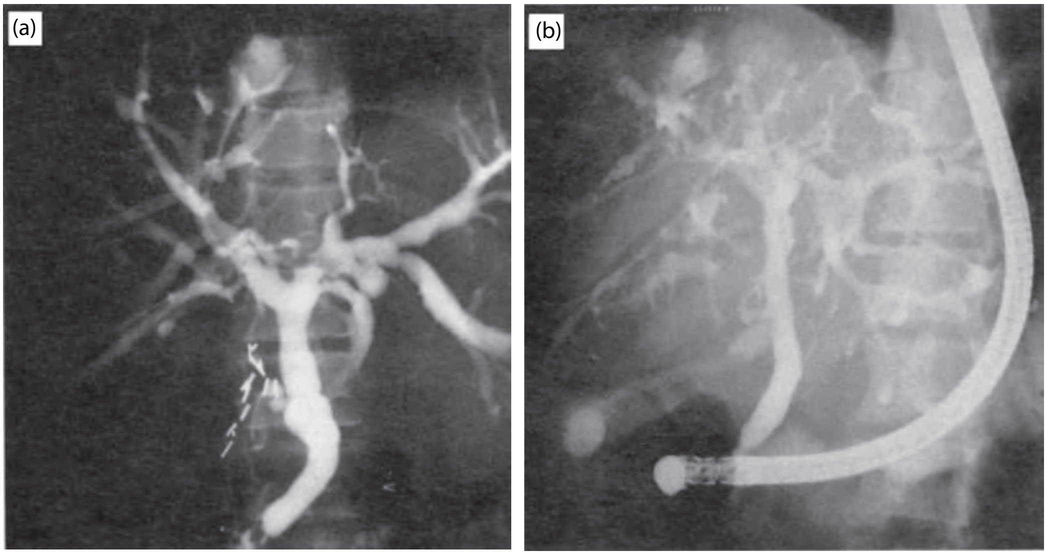


Figure 25.38 Caroli disease. (a) Percutaneous transhepatic cholangiography showing multiple intrahepatic cysts; (b) endoscopic retrograde cholangiopancreatography showing multiple intrahepatic cysts. The patient presented with bleeding from oesophageal varices. Liver biopsy showed hepatic fibrosis. The combination of diffuse intrahepatic cystic disease and hepatic fibrosis is known as Grumbach disease.

cases have been diagnosed during the first year of life. Another 35% become clinically manifest over the next 10 years. The symptoms include cholestatic jaundice, abdominal mass and pain. Complications of the disease include recurrent cholangitis and pancreatitis, hepatic abscess formation, calculous disease, biliary cirrhosis, rupture of the cyst with biliary peritonitis (rare) and portal vein thrombosis. There is also an increased risk of CC, which usually develops in the posteromedial wall of the cyst.

The diagnosis can be established non-invasively with ultrasound and CT scanning. Better definition of the pathological anatomy is, however, obtained with MRCP or ERCP, which also demonstrates the anomalous junction of the bile and pancreatic ducts and its angle (Figures 25.39 and 25.40). Type IV (intraduodenal diverticulum, choledochocele) can be recognized endoscopically as a smooth compressible elevation associated with an enlarged papilla which protrudes into the duodenum.

Treatment

The treatment of choice is surgical excision with a Roux-en-Y anastomosis. This gives much better results than drainage procedures and avoids or minimizes the risk of carcinoma. The excision is, however, difficult as the cyst is often adherent to the other structures of the porta hepatis and is best performed in centres with biliary surgical expertise. The treatment of patients with intrahepatic cystic disease is difficult. When localized, partial liver resection is advisable. It is doubtful whether distal drainage, such as choledochoduodenostomy or transduodenal/endoscopic sphincterotomy, materially influences the course of the disease of the widespread intrahepatic variety and may indeed enhance the risk of cholangitis.

Gallstones

There has been a marked rise in the incidence of gallstones in the West during the past century. The current mean prevalence in Europe obtained from autopsy studies is 18.5% with the lowest prevalence being reported in Ireland (5%) and the highest in Sweden (38%). In the UK, USA and Australia, the prevalence rates vary from 15% to 25%. In every Western country, the prevalence of gallstones in females is approximately twice that of the male population. The highest prevalence is found in the Pima Indian tribe of Arizona with total and female prevalence of 49% and 73%, respectively. A high prevalence rate is also found in South American countries. Gallstones are rare

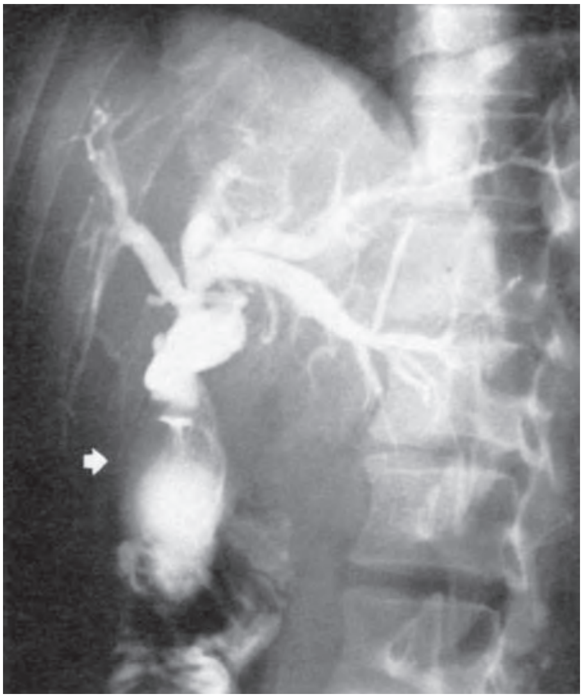


Figure 25.39 Endoscopic retrograde cholangiopancreatography outlining diffuse choledochal dilatation (type I). This is the commonest variety and accounts for 80% of cases.

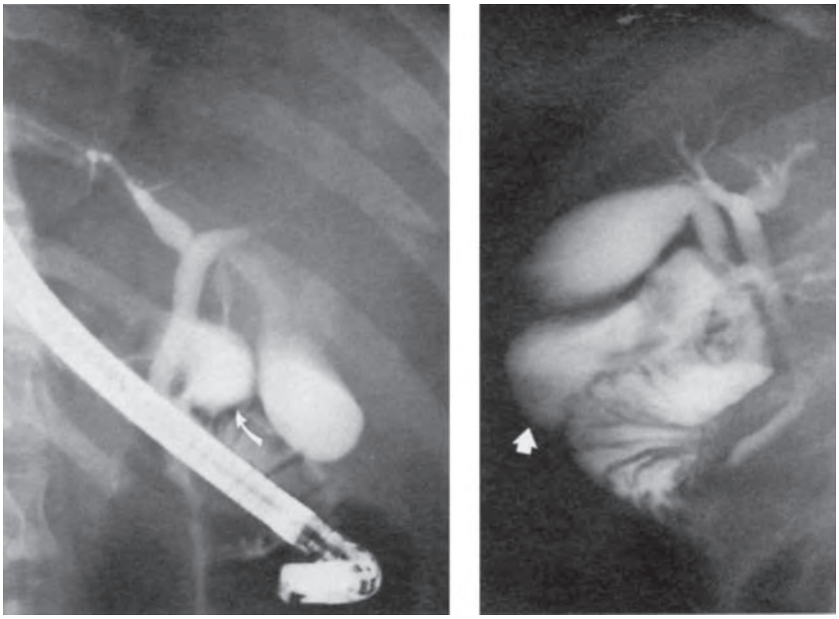


Figure 25.40 Endoscopic retrograde cholangiopancreatography demonstrating a supraduodenal diverticulum (type III) beneath the gallbladder. The arrows point to the diverticulum.

in Africa and most recorded rates in this continent are below 1%. The prevalence of gallstones in Japan has also risen since the early twentieth century from 2% to 7%. In addition, the composition of stones in Japan has changed during the same period such that cholesterol stones now predominate over pigment stones.

Classification

The old classification of Aschoff into inflammatory, metabolic, static and mixed does not correlate with the present pathophysiological information relating to the formation of gallstones and is not currently used. Likewise, the classification into 'pure' and 'mixed' stones is inaccurate as the overwhelming majority of stones are composed of more than one component. The accepted current classification recognizes three main types of gallstones: cholesterol, black pigment and brown pigment stones.

Cholesterol stones

These form in the gallbladder and are preceded by the formation of biliary sludge (see below). They account for the vast majority (75%) of gallstones encountered in the West. They have a protein matrix, and although they are composed predominantly of cholesterol they contain bile pigments and varying amounts of calcium carbonate and palmitate. These calcium salts are deposited on the periphery of the stones and their amounts determine the radiodensity of these yellow gallstones. Cholesterol gallstones do not commonly harbour bacteria (10%) and are not usually associated with infected bile. They are often radiolucent and cast strong acoustic shadows on ultrasonography. Cholesterol stones are often multiple and medium sized, but when solitary they attain a large size and have a radiating crystalline cross-sectional appearance.

Black pigment stones

Black pigment stones form in the gallbladder and account for 25% of stones in the West although their prevalence is higher in Asian countries. They are composed of bilirubin polymers without calcium palmitate, varying amounts of cholesterol (3–25%) and a matrix of organic material. Associated infection is present in less than 20% of patients. Black pigment stones are usually multiple, small, irregular and dark green to black in colour. Although haemolytic states predispose to the formation of black pigment stones, the vast majority occur in patients without detectable chronic haemolysis. Their consistency is hard and they have a layered cut surface.

Brown pigment stones

In contradistinction to the above types, brown pigment stones form in the bile ducts (primary ductal calculi) and are associated with infection of the biliary tract. Scanning electron microscopy of fresh specimens demonstrates bacteria inside crevices and pits of these amorphous soft stones in 98%. Brown pigment stones contain calcium bilirubinate, calcium palmitate and only small amounts of cholesterol bound in a matrix of organic material.

Aetiology

There is an increased prevalence of gallstones in females and the frequency of gallstones increases with age in both sexes. Although familial incidence remains unproven, a positive family history of gallstones is more often obtained from patients with symptomatic gallstone disease than in controls. The importance of genetic and ethnic factors is exemplified by the unusually high prevalence of gallstones in the American Indians, particularly the Pima tribe. The genetic disorder in this ethnic group results in the production of a supersaturated bile and a deficient secretion of bile acids by the liver. Certain risk factors are known to increase the prevalence of gallstones; others induce symptomatic disease in patients with silent gallstones without necessarily enhancing the overall frequency (Table 25.2).

Table 25.2 Risk factors for gallstone prevalence and symptomatic gallstone disease

Increased prevalence	**Precipitation of symptomatic disease**
Female sex*	Pregnancy
Obesity*	Clofibrate
Age*	Thiazide diuretics
Genetics and ethnic factors*	?Oral contraception
Highly refined, fibre-depleted, high animal fat diet*	
Diabetes mellitus*	
Ileal disease and resection	
Haemolytic states*	
Infections of the biliary tract†	
Parasitic infestations†	
Cirrhosis†	
Cystic fibrosis	

*Increased prevalence of cholesterol stones.
†Increased prevalence of pigment stones.

Although it is generally stated that the gallstones which form in patients with ileal disease or after ileal resection are of the cholesterol variety, recent studies have demonstrated that a substantial number of these stones are of the pigment variety. The enhanced lithogenicity and increased incidence of gallstones in the obese are the consequence of an increased hepatic synthesis and secretion of cholesterol. The supersaturation of bile with cholesterol is found in maturity-onset diabetes but not in the juvenile type. The exact mechanism responsible for bile lithogenicity in maturity-onset diabetes is not known. The enhanced incidence of gallstones in children with cystic fibrosis has been attributed to the abnormal mucus that impairs bile flow and favours nucleation.

Early reports demonstrated an increased incidence of symptomatic gallstones in females on the oral contraceptive pill, the relative risk being estimated at 2.5. However, more recent studies have been unable to confirm this finding. This discrepancy has been attributed to the lower oestrogen component of the modern 'mini' contraceptive pill. Studies on the prevalence of gallstones in asymptomatic populations with ultrasound scanning have shown no difference between women on the contraceptive pill and those who are not. Some reports have indicated that the modern 'mini' contraceptive pill may be associated with an increased risk of symptomatic disease in young women (aged 29 years or less).

Pathogenesis of gallstones

Our knowledge of the pathogenesis of gallstone formation remains incomplete, although there has been considerable progress on the mechanisms involved during the past two decades. Undoubtedly the pathogenesis of cholesterol stones is different from that of black pigment stones; and brown pigment stone formation differs from both.

Cholesterol stones

The formation of cholesterol stones involves seven processes:

- supersaturation of bile with cholesterol
- incomplete transfer of cholesterol from the biliary vesicles to the bile salt micelles and formation of abnormal high cholesterol-containing biliary vesicles
- aggregation and fusion of unstable vesicles
- cholesterol crystallization: nucleating and antinucleating factors
- biliary sludge formation
- stone growth.

Supersaturation with cholesterol

The outstanding biochemical abnormality associated with the formation of cholesterol gallstones is the secretion of bile that is supersaturated with cholesterol. Under normal physiological conditions, cholesterol is secreted by the hepatocytes into the hepatic bile as cholesterol–phospholipid vesicles. Within the bile a phase change occurs due to the relative high concentrations of bile acids, and micelles form. These are essentially molecular aggregates in which the cholesterol and phospholipid molecules form a central core surrounded by bile salt molecules. This mechanism is invoked to maintain cholesterol (which is water insoluble) in solution. Within the narrow range of water content of bile (80–95%), the solubility relationships of the biliary lipids can be expressed by plotting their relative molar concentrations on triangular coordinates (Figure 25.41). In the micellar zone, which represents the normal, the cholesterol is in solution as micelles and the bile is unsaturated with respect to cholesterol. Thus cholesterol precipitation and stone formation do not occur. Other methods are used to express the solubility of cholesterol in bile. These include the percentage saturation, the lithogenic index and the ratio of the concentrations of bile salts and phospholipids over the concentration of cholesterol.

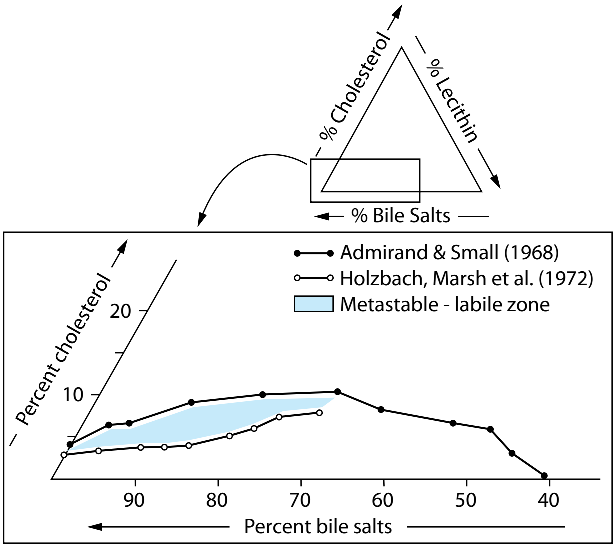


Figure 25.41 Solubility relationships of cholesterol, bile salts and lecithin (phospholipid) expressed by plotting their relative molar concentrations on triangular coordinates. Points below the metastable–labile zone indicate unsaturated bile. (From Popper, Schaffner (eds) *Progress in Liver Disease*, vol. V. New York, NY: Grune and Stratton, 1976.)

The lithogenic index is the ratio of the actual amount of cholesterol that can be dissolved in the bile sample using the triangular coordinate plots. A lithogenic index of unity or greater indicates that the bile is supersaturated with respect to cholesterol. The normal ratio is concentration bile salts + concentration phospholipids/concentration cholesterol = 10:1.

The source of the supersaturated bile is the liver, from either increased synthesis of cholesterol or a decreased synthesis/secretion of bile salts and phospholipids. Excess cholesterol secretion in the bile is well documented in obese patients and is thought to result from the increased activity of the enzyme hydroxymethylglutaryl coenzyme A (HMG CoA) reductase resulting from the chronically elevated levels of insulin found in overweight individuals. Most non-obese patients with cholesterol gallstones have a reduced absorption as the cause of the diminished bile salt pool in these patients. The most likely explanation of the smaller pool (and supersaturated bile) is an exaggeration of the normal feedback mechanism between the bile salt return and the hepatic synthesis of bile acids together with an increased enterohepatic cycling possibly consequent on abnormal excessive gallbladder contractility. Although gallstones develop only in individuals with supersaturated bile, many patients with this abnormality never develop gallstones. Thus other factors are involved. These include the persistence of biliary vesicles and kinetic factors.

Biliary vesicles

Under normal conditions, the cholesterol–phospholipid vesicles are relatively stable and disappear as micelles form and take up their cholesterol and phospholipid constituents. In the pathological state, the phase change to micelles is incomplete and as more phospholipid than cholesterol is extracted into the micellar aggregates with bile salts, biliary vesicles with an abnormally high cholesterol–phospholipid ratio are produced. These abnormal vesicles are unstable and are the source of cholesterol monohydrate crystals in the bile. The crystallization is preceded by aggregation and fusion of the high cholesterol-containing vesicles.

Kinetic balance between nucleating and antinucleating factors

Bile contains substances which either inhibit (apolipoprotein A1) or promote (mucin) the growth of cholesterol crystals. Thus the development of cholesterol gallstones is influenced by the balance between these two kinetic factors. An anionic polypeptide fragment (APF) of high-density lipoprotein occurs in bile and together with immunoglobulin A constitutes the lipoprotein complex of bile. It is thought that an absolute or relative decrease in the amount of APF favours nucleation and growth of cholesterol monohydrate and pigment crystals.

The biliary proteins include enzymes, transport proteins, hormones, plasma proteins and mucin. The evidence for the important role of mucin in the nucleation and growth of cholesterol crystals is overwhelming. Hypersecretion of densely glycosylated mucin is observed in experimental models of gallstone formation; mucin is a major component of biliary sludge and stone formation appears to start in the mucous gel. Thus gallbladder mucus is thought to be involved both as a nucleation-promoting agent and in providing the right milieu which encourages stone growth. Recently a protein with marked nucleation promotion activity for cholesterol crystals that binds to concavalin A has been isolated from the bile of patients with and without gallstones. Its activity was found to be increased in patients harbouring gallstones.

Biliary sludge

Biliary sludge is composed of mucin, calcium, monoconjugated bilirubin and cholesterol and is now thought to be the direct precursor of gallstones. A number of reports have documented symptoms including biliary-type pain in patients who were found to have biliary sludge but no gallstones.

Other factors

There is no doubt that the above account presents an incomplete picture of the pathogenesis of cholesterol stones. The role of calcium is indicated by the presence of calcium salts in the majority of stones, and the experimental demonstration in the prairie dog model that cholesterol vesicle aggregation and the induction of cholesterol gallstones is associated with an increase in the total and free ionized calcium concentrations in the gallbladder bile.

That the gallbladder plays an important role in gallstone formation (both cholesterol and black pigment) is evident by the fact that 85–90% of stones are encountered in this organ rather than in the bile ducts. It has been postulated that the gallbladder may alter the physicochemical composition of bile, favouring nucleation and crystal growth by abnormal absorption/secretion, defective surface pH, stasis resulting from impaired gallbladder emptying and stratification of bile (Figure 25.42) or by providing essential nucleating factors including mucin, desquamated cells, bacteria and refluxed intestinal contents. It is probable that gallstones are formed at different occasions depending on the balance between nucleation inhibition and promoting factors at different time periods (Figure 25.43).

Black pigment stones

Much less is known about the pathophysiology of black pigment stones. In cirrhotic patients who have a high prevalence of these stones, an elevated concentration of monoconjugated bilirubin and a lower bile salt concentration than normal has been documented. The hypothesis that patients with black pigment stones have a bile which is supersaturated with calcium bilirubinate has had some experimental backing in a dietary model of pigment stones by which a significant increase in the gallbladder concentrations of unconjugated bilirubin and calcium were documented. Calcium is a universal component of black pigment stones and both free and total ionized calcium is increased in canine models of pigment gallstone disease. As the hepatic bile concentration is unchanged, the altered chemical composition is thought to be secondary to altered gallbladder function.

Patients with bile acid malabsorption (after ileal resection) have low biliary cholesterol in addition to a reduced bile salt pool. They exhibit an increased risk of gallstone formation, predominantly of the black pigment kind. This observation has led to the hypothesis that biliary cholesterol may exert a protective action on the gallbladder mucosa and, accordingly, low biliary cholesterol renders the mucosa susceptible to direct

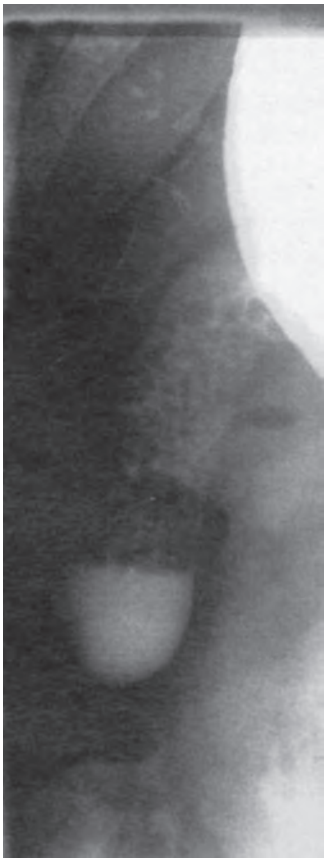


Figure 25.42 Stratification of bile with floating gallstone layer.

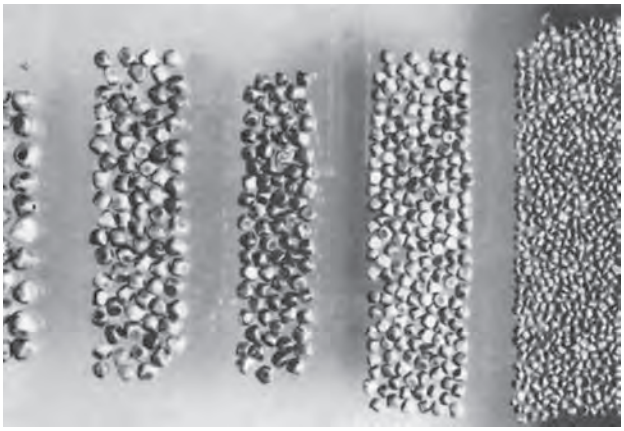


Figure 25.43 Several generations of cholesterol stones removed from a gallbladder after cholecystectomy. The factors within the gallbladder which favour precipitation of cholesterol crystals and the growth of cholesterol microliths into discernible stones are outlined in the text.

injury by bile acids as these are concentrated in the gallbladder. There is no proof for this hypothesis.

There is increasing evidence for the important role of some mucins (especially MUC3 and MUC5B) in the development of pigment stones, especially those forming in the intrahepatic ducts. The peribiliary glands are the source of the mucins. A recent study has demonstrated an increased mRNA expression of MUC3 and MUC5 in the cells of the peribiliary glands of stone-harbouring intrahepatic ducts compared to normal controls.

Brown pigment stones

These ductal calculi are caused by infection by Gram-negative bacteria such as *E. coli* and *Bacteroides fragilis*, which elaborate and release β-glucuronidase in the bile. Recent studies have documented infection in 80–98% of cases. In many Eastern countries where infestation with *Ascaris lumbricoides* is endemic, the eggs of this parasite have been repeatedly identified in the nucleus of brown pigment stones, but the cause is the associated bacterial infection of the biliary tract. Likewise, the formation of both intrahepatic and extrahepatic ductal calculi in recurrent pyogenic cholangitis is due to infection of the biliary tract by Gram-negative bacteria. The bacterial β-glucuronidase is implicated in the hydrolysis of conjugated bilirubin with consequent precipitation of insoluble calcium bilirubinate. Brown pigment stones are encountered in biliary tract conditions associated with stasis and infection such as chronic obstructive disease, indwelling biliary endoprostheses and around non-absorbable suture material or metal clips used in biliary tract surgery.

Clinical syndromes of gallstone disease

The symptomatology of gallstone disease is varied. Often non-specific, the symptoms may be acute, chronic or totally absent when gallstones are diagnosed as an incidental finding during the investigation of patients for unrelated disorders. The differentiation between silent and symptomatic gallstones is important since this affects management in the individual case.

In patients with chronic symptoms, it is important to stress that the demonstration of gallbladder disease by oral cholecystography/ultrasound scanning does not exclude other disorders being responsible for the symptoms, and a careful clinical evaluation, together with the appropriate investigative protocol, is essential in all patients with chronic symptoms and ultrasonically confirmed gallstone disease. This is especially important in the selection of patients for elective cholecystectomy. The common coexisting diseases include:

- colonic motility disorders and diverticular disease
- and peptic ulceration
- reflux oesophagitis and hiatal hernia
- pancreatitis
- colonic cancer
- renal disease
- ischaemic heart disease.

In addition to gallbladder imaging, an upper gastrointestinal endoscopy or barium series and, in certain situations, a barium enema is advisable in all patients undergoing elective cholecystectomy for chronic symptoms.

Symptomless (silent) gallstones

Most surveys have shown that silent gallstones heavily outnumber the symptomatic ones. Silent gallstones are diagnosed as incidental findings most commonly by abdominal radiographs. The previous controversy regarding the management of asymptomatic gallstones has been resolved by prospective studies which have shown that the vast majority of silent gallstones will not cause symptoms or complications during life.

Comparative evaluation of expectant vs surgical management of asymptomatic gallstones has shown that cholecystectomy reduces marginally the life expectancy in addition to being substantially more costly. Another argument for cholecystectomy in the past has been the prevention of gallbladder cancer, the development of which is known to be associated with the presence of gallstones. However, carcinoma of the gallbladder is rare and the overall operative mortality with the widespread adoption of prophylactic cholecystectomy in patients with silent gallstones would certainly exceed that due to cancer of the gallbladder by a significant margin. The evidence linking cholecystectomy with the development of colon cancer remains conflicting and cannot be used as a further argument against prophylactic cholecystectomy. The consensus of current surgical opinion is that there is no indication for cholecystectomy in the management of patients with asymptomatic gallstone disease.

There are four exceptions to this policy. The first relates to patients undergoing surgical intervention for other conditions in which gallstones are found at laparotomy. In one report, 50% of these patients developed complications or symptomatic disease subsequently and 12% required cholecystectomy within 30 days after the operation. On the other hand, concomitant cholecystectomy increases the postoperative morbidity. Thus if incidental gallstones are found at operation, cholecystectomy is indicated if the patient's general condition is good (ASA I, II). The second group which merits cholecystectomy for asymptomatic gallstones is made up of acromegalic patients on long-term treatment with somatostatin analogue, which often produces large gallstones. As somatostatin treatment has to be continued indefinitely, LC is indicated in these patients, once stones are documented by ultrasound. Presently, prophylactic removal of the gallbladder (prior to stone formation) is not practised in these patients. Patients with a calcified gallbladder have a significant risk of developing cancer of the gallbladder and for this reason cholecystectomy is indicated even if the condition is asymptomatic. Diabetic patients with gallstones are more prone to develop symptoms and complications and for this reason some advocate LC in this group although this remains debatable.

Symptomatic gallstone disease

The clinical presentation varies and may be acute or chronic.

The spectrum of symptomatic gallstone disease is:

- chronic cholecystitis
- acute biliary colic/acute cholecystitis
- jaundice due to large bile duct obstruction
- cholangitis/septicaemia
- acute gallstone pancreatitis
- biliary fistulous disease
- gallstone ileus.

Chronic cholecystitis

Chronic inflammation of the gallbladder is most commonly due to stones and the term 'chronic cholecystitis' should be restricted to gallbladders containing gallstones with varying degrees of inflammation, from mild mucosal/submucosal changes to gross transmural fibrosis leading to a contracted fibrous encasement of the biliary calculi.

Symptoms and signs

Most commonly, patients with chronic cholecystitis complain of recurrent attacks of epigastric or right hypochondrial pain, often radiating to the right side of the back and, less commonly, to the shoulder blade. The pain is more often persistent than intermittent. Episodes of biliary colic with severe intermittent peaks of pain lasting a few minutes to several hours may subside spontaneously or progress to cystic duct obstruction and acute cholecystitis. Nausea and vomiting may accompany both episodes of persistent pain and severe attacks of biliary colic. Jaundice and dark urine may follow an attack and indicate common bile duct obstruction by a calculus. The jaundice often subsides after a few days but may persist as a major presenting symptom. It is now established that indigestion, dyspepsia, flatulence, intolerance to fatty foods, abdominal distension and belching occur with the same frequency in the general population as they do in patients with chronic gallstone disease.

The only reliable sign, which is infrequently found on clinical examination, is tenderness in the right upper quadrant. More often than not, the clinical features of chronic cholecystitis are non-specific and confirmation by imaging tests (usually ultrasonography) is essential for diagnosis in the vast majority of cases.

Treatment

Cholecystectomy

The treatment of chronic cholecystitis is surgical – cholecystectomy. There is little doubt that these patients should have their gallbladder removed, as approximately 30% of them will develop complications if surgical treatment is delayed. Furthermore, the morbidity and mortality following surgical intervention are enhanced in those patients who develop complications necessitating surgical intervention. In practice, the problem concerns the selection of patients, which should be based on establishing that the gallstones are the cause of the patient's symptoms since these are by no means specific and can be caused by other common gastrointestinal disorders. Poor case selection accounts for a large cohort of those patients who continue to experience symptoms after cholecystectomy (postcholecystectomy syndrome). Some of these patients are subsequently found to have disease outside the biliary tract.

Peroperative cholangiography should be considered as an integral part of cholecystectomy. The cholangiographic findings, the presence of jaundice together with the operative appearances, dictate the need for exploration of the common bile duct. Exploration of the common bile duct is usually followed by the insertion of a T-tube even in those patients who are found to have a negative exploration, although some surgeons would opt not to insert a tube in patients with a negative exploration and simply close the choledochotomy. Others insert a small cannula through the cystic duct remnant to enable the performance of postoperative cholangiography before discharge from hospital. The T-tube should not be smaller than 14 Fr and its long limb should be brought out by the shortest route laterally in the right subcostal region. These measures facilitate considerably the removal of retained ductal stones through the T-tube tract by means of the flexible choledochoscope (see below). The Whelan–Moss T-tube was specially designed to create a wide tract between the bile duct and the abdominal parietes and thus facilitate percutaneous stone extraction via the T-tube tract. Its long limb is wider than the intracholedochal portion.

Opinion on the need for drainage of the gallbladder fossa remains divided. The argument against drainage has been strengthened by the results of several prospective clinical trials since these have either failed to show a difference between the drained and the undrained groups or indicated that drainage increases the postoperative sepsis and morbidity. It seems likely that the routine uneventful cholecystectomy without exploration of the common bile duct does not require the insertion of a subhepatic drain, but most biliary surgeons would still recommend drainage for the following:

- difficult cholecystectomy
- early cholecystectomy for acute cholecystitis
- in patients who require exploration of the common bile duct.

The mortality of elective open cholecystectomy for chronic cholecystitis is low (0.3–1.0%) but is higher in the elderly (5–6%). LC is now firmly established as the gold standard therapy for symptomatic gallstone disease. With experience and training, it is applicable to over 95% of patients. The procedure reproduces all the steps of the traditional open cholecystectomy, which it has largely replaced because of its well-documented advantages. Conversion to open surgery is indicated if a safe dissection of the structures in the triangle of Calot cannot be performed because of dense fibrosis, severe adhesions from previous surgery, the presence of a Mirizzi syndrome and severe acute disease with gross inflammatory oedema. Conversion to open cholecystectomy may also be necessitated because of the onset of a complication (severe bleeding, bile duct injury, etc.) which cannot be safely dealt with by the laparoscopic approach.

The advantages of LC over the conventional operation include less postoperative pain, virtual absence of ileus, short hospital stay (0–2 days) and accelerated return to full activity or work (within 10–14 days). In addition, there is an overall reduction of the incidence of postoperative wound and chest infections, wound dehiscence, incisional hernia and deep vein thrombosis. The postoperative mortality appears to be lower than that of open cholecystectomy. Other advantages include reduced adhesion formation and wound complications. The downside is an increased incidence of iatrogenic bowel, vascular and biliary injuries. The importance of routine IOFC during LC must be stressed. The overriding reason for routine IOFC is the provision of a road map of the biliary anatomy and the determination of a safe site for medial ligature or clipping of the cystic duct without compromise of the extrahepatic bile conduit. The other reasons for routine IOFC are similar to those for open cholecystectomy: detection of biliary tract anomalies and unsuspected ductal calculi. If the surgeon is experienced, these calculi are best removed during the operation either by extraction through the cystic duct or by direct laparoscopic common bile duct exploration.

Single port laparoscopic cholecystectomy

A recent modification is a single port LC in which a single port (disposable or reusable) is used that enables the insertion of a 5 mm laparoscope and at least two instruments (Figure 25.44).

However some surgeons prefer to use a single incision (subumbilical) and insert the three separate 5.0 mm ports and instrument without use of any special single multi-instrument port (*single incision technique*). Single port or single incision LC has been shown to reduce postoperative pain and improve cosmesis as the scar is 3.0 cm and is located in or just below the umbilicus. It is however more difficult than the conventional multiport LC, although the reported data indicate equivalent safety between the two approaches.

To date, a broad spectrum of operations have been performed by the various specialties by the single port laparoscopic (SPLS) approach: cholecystectomy, colorectal resection, totally extra-peritoneal inguinal hernia repair**,** bariatric surgery including sleeve gastrectomy, gastric banding**,** gastric bypass, fundoplication, splenectomy, laparoscopic urology, e.g. nephrectomy, pyeloplasty, and prostatectomy, laparoscopic gynaecology, e.g. salpingostomy, salpingectomy, cystectomy and hysterectomy. Worldwide SPLS cholecystectomy has been the commonest operation performed in the published literatures with more than 2000 cases reported to date.

The safety and efficacy of the SPLS for LC has been confirmed by randomized and case cohort studies. One prospective randomized controlled trial of traditional multiport LC vs SPLS confirmed equivalent safety of the two approaches. In this trial, the cosmetic score was higher for SPLS than with traditional LC. However, despite the enthusiasm of many surgeons in adopting the technique, others remain unconvinced of the real benefit for the patients, especially with reference to reduction of the postoperative pain and adhesion formation and furthermore have concerns on the increased risk of periumbilical hernia formation with SPLS.

Compared with the traditional multiport laparoscopic surgery, there are a number of major constraints in ergonomics of and technologies for SPLS, which have hindered the wide spread uptake of this approach. Even with the best operating set-up, image display system, port and instrumentation available, the SPLS approach imposes several restrictions: maintenance of sufficient exposure, sustained pneumoperitoneum, adequate retraction, collision between instruments (internal and external), collision between instruments and optics, and limited instrument manipulation and triangulation. For this reason, many surgeons have had to overcome these restrictions by various techniques such as using percutaneous sutures for retraction and employing coaxial, flexible, and articulating instruments to improve triangulation.

Different operating systems, various ports, types of instruments, and endoscopes have been tried and used which have, to some extent, reduced the constraints imposed by the SPLS approach. Undoubtedly further more advanced instrumentation such as hand-held manipulators with 7 degrees of freedom and robotic devices for SPLS are needed to realize the full potential of the SPLS approach.

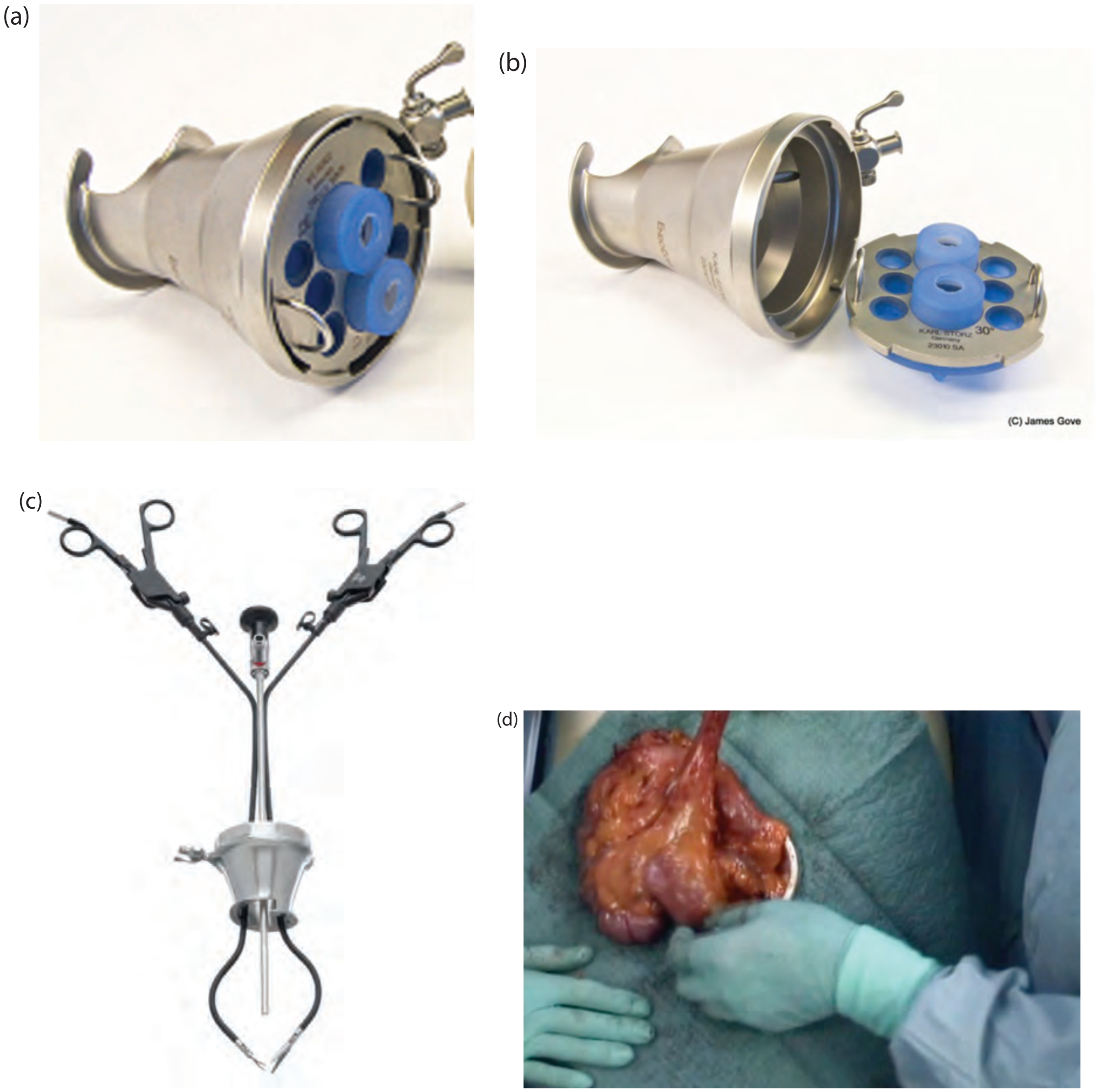


Figure 25.44 Dundee EndoCone for single port laparoscopy (Storz, Tuttlingen). This reusable port is introduced by a screwing movement after creation of a 3.0 cm transumbilical incision and is self-retaining. The detachable bulkhead (cap) has two central large valves (admitting instruments up to 13 mm including staplers) and six 5.5 mm valved ports for 5 mm instruments which are specially designed: they have distal curves to restore triangulation and proximally diverting handles that permit separation of the two hands of the operator, essential for ergonomic manipulations.

Mini-cholecystectomy

This implies performance of cholecystectomy through a small (5 cm wound) that should be placed in the midline and not in the right subcostal region. Proponents recommend this technique for both elective cholecystectomy and patients with acute cholecystitis. The procedure uses standard operating techniques with certain modifications such as the use of a headlight, a ring retractor, clip applicators and long instruments. The need to extend the incision is encountered in 15% of patients. A more recent modification has been described as ‘cylindrical cholecystectomy’. The operation is based on the introduction of a 3.8–5.0 cm diameter cylinder that is 10 cm long which isolates the hepatocystic region from the surrounding structures, and thus facilitates the intervention.

In general, mini-cholecystectomy is associated with a shorter hospital stay (3.5 days vs 8.5–11 days) than patients undergoing cholecystectomy through an unrestricted incision, and some have reported an earlier return to work after the procedure.

The procedure can be technically difficult especially in patients with excessive fat in the triangle of Calot and most surgeons have abandoned it in favour of the laparoscopic approach.

Partial cholecystectomy

This technique was first described by the group at Cape Town for cholecystectomy in patients with cirrhosis. Many variations of this technique have been described involving a subtotal (fundus first) resection of the gallbladder but leaving the posterior wall attached to the hepatic bed. The Cape Town group reported on this technique performed by the laparoscopic approach for patients with complicated acute cholecystitis or fibrosis and encountered one postoperative death (myocardial infarction) and three bile leaks in 29 patients. Thus suction drainage of the gallbladder bed is necessary after subtotal cholecystectomy (open or laparoscopic).

A modification using a 1 cm rim of Hartmann's pouch to buttress and occlude the internal opening of the cystic duct and leaving the structures of Calot's triangle undisturbed was described by Schein, who reported favourably on the technique in 16 elderly/high-risk patients. This author emphasizes its speed (mean operating time of 40 minutes), technical safety in obviating a difficult dissection of the liver bed and the practical advantage compared with other conservative surgical procedures that leave the gallbladder *in situ,* of preventing gallstone formation.

Cholecystolithotomy

Patients with previous vagotomy for ulcer disease frequently develop symptomatic gallstones. Cholecystectomy in vagotomized patients with a functioning gallbladder incurs the risk of severe explosive diarrhoea in a high percentage of cases. For this reason, cholecystectomy is not warranted in this subset of patients and removal of the gallstones (cholecystolithotomy) constitutes the appropriate treatment. This may be followed by maintenance therapy with oral bile salt in an attempt to prevent recurrence if the stones are of the cholesterol variety (see below). The best approach is laparoscopic cholecystolithotomy with closure of the gallbladder with absorbable suture after gallstone clearance has been achieved. Others favour radiological guided stone extraction, fragmentation or dissolution. A cholecystectomy is the right treatment in vagotomized patients if the gallbladder is non-functioning.

Dissolution by MTBE and percutaneous lithotripsy

Methyl tert-butyl ether (MTBE) is a powerful organic solvent which dissolves cholesterol gallstones within hours. It is administered locally via a pigtail or other self-retaining catheter introduced percutaneously (usually through the hepatic substance) into the gallbladder under radiological control. MTBE is then instilled and left in the gallbladder for several hours. When dissolution is achieved, the gallbladder liquid contents are aspirated and the catheter is removed. MTBE is toxic following systemic absorption and can cause local damage if it escapes into the common duct and duodenum (erosions). For this reason, microprocessor-controlled pump delivery systems (pressure and volume controlled) are used to obviate spillage of the solvent into the ductal system. MTBE dissolution is rarely used nowadays and is indicated as one of the alternatives when removal of the gallbladder is contraindicated (poor-risk patients, previous vagotomy) or the patient specifically requests removal of the stones but not the gallbladder.

Alternatively, a rotary mechanical lithotriptor (Kinsey–Nash) can be inserted percutaneously under radiological control into the gallbladder. This device operates by creating a vortex that draws stones to a high-speed rotating impeller that fragments the calculi. Although effective, this technique results in damage to the mucosa by the bombardment with stone fragments and the limited reported literature on its use indicates a high incidence of complications. This method has been abandoned.

Extracorporeal shockwave lithotripsy

With these devices, extracorporeal shockwaves produced by spark-gap (Dornier system), piezoceramic (Wolf system) or electromagnetic (Siemens system) generators are focused by a concave reflector and targeted under ultrasound guidance to the stones, thereby inducing fragmentation. The third-generation machines avoid the need for immersion in a water bath and general anaesthesia. Although extracorporeal shockwave lithotripsy (ESWL) can achieve successful stone fragmentation in 80% of patients with solitary small gallstones at 1 year, it often fails if the stones are large (>3 cm), multiple and calcified. It is thus applicable to only 19% of patients with symptomatic gallstones. Furthermore, repeated treatment is often necessary to achieve complete fragmentation and this has to be followed by maintenance therapy with oral bile salts, despite which stone recurrence occurs in 50% at 5 years. Compared with surgical treatment, ESWL is costly, unreliable and has a limited applicability. Contrary to its established value in the treatment of renal calculi, ESWL is seldom used in the treatment of gallstones and its current role seems restricted to fragmentation of occluding ductal calculi in jaundiced patients. Even this indication is not established and remains to be defined in large patient studies. Currently, fragmentation of ductal calculi by ESWL can be achieved in 70–90% of patients with an average of two lithotripsy sessions. However, some 50% of patients require additional procedures (endoscopic extraction) to achieve stone removal, including operative intervention in 10%. Complications such as haematuria, biliary pain, sepsis and haemobilia are encountered in 15% of patients.

Oral bile salt therapy

Oral dissolution is applicable only to patients with cholesterol stones in a functioning gallbladder. The administration of the primary bile acid chenodeoxycholic acid and its 7-β epimer, ursodeoxycholic acid, reduces the cholesterol saturation of bile and prolongs nucleation time, thereby resulting in stone dissolution after several weeks/months of therapy. Following successful dissolution, maintenance therapy with low-dose ursodeoxycholic acid is required. Despite this, the risk of gallstone recurrence is 12.5% after the first year and 61% by the eleventh year. Oral dissolution fails if the gallstone load is large (larger than 3 cm or multiple) and if the stones are calcified. The disadvantages of this treatment of gallstones include high cost, high recurrence rate and limited applicability. Currently, the place of bile salt therapy is restricted to patients in whom cholecystectomy is contraindicated, either as primary treatment or after gallstone extraction, fragmentation or dissolution by MTBE. Another important indication of oral bile salt therapy is in patients with endoscopic stents as a prophylactic measure to reduce the incidence and severity of stent encrustation with calcium bilirubinate.

Chemical sclerosis

Experimental sclerosis of the gallbladder (chemical cholecystectomy) after occlusion of the cystic duct has been achieved by instillation of various sclerosants in various animal models. However, none of the agents used to date is safe enough to use clinically. The other problems with this approach include regeneration of the gallbladder mucosa from the cystic duct remnant and the possible late risk of malignant change. Physical methods of destruction of the gallbladder mucosa are currently being studied in animal experiments. These include photodynamic laser treatment.

Acute biliary colic and acute cholecystitis

Acute biliary colic results in severe colicky abdominal pain usually accompanied by nausea and vomiting. The duration of the severe pain, which makes the patient restless, varies from 30 minutes to several hours. Biliary colic often merges into acute obstructive cholecystitis. However, resolution of the severe colicky painful episode either spontaneously or as a result of analgesic medication/antispasmodic without the development of acute cholecystitis is common, and many patients give a history of recurrent episodes of biliary colic before the development of acute cholecystitis.

Between 10% and 30% of patients undergoing cholecystectomy present with acute cholecystitis. This is most commonly obstructive in nature (>95%) from impaction of a stone in Hartmann's pouch/cystic duct. Much less commonly, acute cholecystitis is acalculous, although the incidence of this complication in critically ill patients is diagnosed more frequently nowadays. Acute cholecystitis in elderly people may result from cystic duct obstruction due to carcinoma of the gallbladder. Whereas patients harbouring residual infection of the gallbladder with *Salmonella typhi* as typhoid carriers may cause sporadic outbursts of typhoid, acute salmonella cholecystitis is very rarely encountered in the West.

Acute obstructive (calculous) cholecystitis

Pathology

The attack develops when the cystic duct becomes obstructed by a gallstone impacting in Hartmann's pouch. Following obstruction of the cystic duct/Hartmann's pouch, the gallbladder becomes hyperaemic, oedematous, tense and distended. The initial inflammation is chemically induced and is not of bacterial origin although sepsis is an important feature of the established disease and its complications. Cultures of gallbladder bile taken during open cholecystectomy are positive in only 15–30% of cases and in only 3% of patients undergoing LC for the same disease. The predominant micro-organisms isolated from the gallbladder bile in these patients are *E. coli* (60%), *Klebsiella* spp. (22%) and *Streptococcus* spp. (18%). There appears to be no relation between positive gallbladder cultures and postoperative wound infections for both open cholecystectomy and LC but the overall incidence of wound infection is much higher after the open procedure (14% vs 5%) and serious wound infections after LC are very rare. Thus the small incisions used in laparoscopic gallbladder surgery may be less susceptible to infective complications.

These observations indicate that the initial inflammatory process following obstruction of the cystic duct is of a chemical nature with infection supervening in some patients during the later stages of the disease. It is believed that trauma, secondary to gallstone impaction, leads to mucosal damage through the release of phospholipases that convert lecithin (a mucosal protective factor against bile acids) to lysolecithin, a known mucosal toxin. Alternatively, the release of the prostaglandin precursor arachidonic acid by the action of phospholipase A on lecithin may mediate the inflammatory response by producing prostaglandins. In the first few days, the bile appears macroscopically normal and is sterile, but, with the progress of the inflammation, absorption of pigments and bile salts takes place, and the contents then vary from a thin mucoid material to frank pus. The histological changes of the established condition involve the mucosa, fibromuscular wall and serosa and vary from mild acute inflammation with transmural oedema to severe disease with patches of necrosis usually in the fundal region of the gallbladder, which becomes wrapped by the greater omentum.

With conservative treatment, the inflammation resolves in some 80% of patients as the rising tension in the gallbladder lumen from the outpouring of the inflammatory exudate lifts the walls of Hartmann's pouch off the impacting stone. When this disengages and drops into the gallbladder lumen, cystic duct drainage leads to resolution. This fortuitous sequence is not encountered in 20% of patients, usually elderly, in whom patchy gangrene and/or perforation with a large inflammatory phlegmon or peritonitis supervene.

The subsequent sequence of events following the acute inflammatory process may vary:

- resolution (most common) with scarring, abnormal function or non-function of the gallbladder
- persistence of the infection: the gallbladder becomes distended with pus (empyema of the gallbladder)
- resolution of the inflammatory process within the gallbladder with persistence of the cystic duct obstruction – mucocele (hydrops) of the gallbladder
- gangrene and acute perforation leading to localized (pericholecystic) abscess or frank biliary peritonitis
- chronic perforation with the development of bilioenteric and biliobilial fistulas.

Whereas the majority of patients diagnosed as acute cholecystitis have the classic acute cystic duct obstruction and its associated inflammatory condition of the gallbladder, others are instances of chronic cholecystitis presenting with acute pain or biliary colic. Although acute cholecystitis is often suspected on clinical grounds, a definitive diagnosis can only be obtained by specific investigations (EHIDA scintiscanning, ultrasonography). The serum amylase should always be performed in addition to the liver function tests. Scout abdominal plain film and chest radiographs are used to exclude perforation and the presence of gas in the biliary tract.

Symptoms and signs

The clinical picture varies with the severity of the inflammatory process. Known pre-existing gallbladder disease may be present or chronic symptoms over several months to years may precede the acute presentation. Alternatively, acute obstructive cholecystitis may be the first intimation of gallstone disease.

In mild cases, the patient complains of right upper quadrant pain and tenderness. Pyrexia, severe pain and tenderness in the right hypochondrium with rebound reflect more severe degrees of gallbladder inflammation. In these instances, Murphy's sign (inspiratory arrest due to pain on inspiration during gentle palpation of the right subcostal region) is usually present. Nausea, vomiting, ileus, mild abdominal distension and toxicity are encountered in the severe forms of the disease. Jaundice is present in 20–25% of patients with acute obstructive cholecystitis but common duct stones are found in only 12% of these patients. In the absence of ductal calculi, jaundice has been ascribed to reactive hepatitis or oedema of the common bile duct. A tender palpable mass in the right subcostal region is found in 25% of cases and signifies one of the following:

- empyema of the gallbladder
- omental phlegmon
- abscess due to localized perforation
- carcinoma of the gallbladder, especially if the patient is elderly.

Laboratory tests are frequently non-specific, their greatest value being to rule out other important conditions in the differential diagnosis, particularly acute pancreatitis. Most patients will have a neutrophil leucocytosis ($>10 \times 10^9$/L) together with some abnormality of the liver function profile. The levels of serum bilirubin and alkaline phosphatases do not invariably correlate with the presence of ductal calculi but are suggestive and clinical jaundice warrants investigation with endoscopic retrograde cholangiography (ERC) or MRCP. Other laboratory findings include raised transaminases and minor elevations of the serum amylase, below the diagnostic threshold for acute pancreatitis.

Differential diagnosis

Usually the diagnosis of acute cholecystitis is not difficult, but other common intra-abdominal conditions, e.g. perforated peptic ulcer, acute pancreatitis or a retrocaecal appendicitis associated with a high caecum and viral hepatitis, need to be considered in the individual case. Enterally transmitted non-A, non-B viral hepatitis can simulate acute cholecystitis quite closely as may right-sided pyelonephritis, lobar pneumonia and myocardial infarction. Aside from routine chest radiography and plain abdominal films, an electrocardiogram is advisable in elderly patients and in those patients with a known history of ischaemic heart disease.

Imaging tests

The yield from a plain abdominal radiograph, although limited, may be detection of important calcified gallstones in 10–20% of patients. Gas in the gallbladder lumen and biliary tract caused by emphysematous cholecystitis (see below) is encountered infrequently but is obviously very important. Absence of free air under the diaphragm is useful for excluding perforated ulcer.

Real-time ultrasonography and biliary scintiscanning form the mainstays in the confirmatory diagnosis of acute cholecystitis. Greyscale B-mode ultrasound is the most commonly used test since it is readily available, non-invasive, quick and easy to perform. Furthermore, it has the advantage of providing information about the liver, biliary tract and pancreas together with other sources of non-biliary right upper quadrant pain. The sonographic features include a positive Murphy's sign, calculi or sludge, thickened gallbladder wall and pericholecystic oedema. The examination is hampered by obesity and overlying bowel gas and is, of course, observer dependent. There is now good evidence from reported studies that the accuracy of ultrasound for the diagnosis of acute cholecystitis is considerably improved with colour velocity imaging and especially with power Doppler when compared with greyscale imaging (sensitivity 95% vs 86%, accuracy 99% vs 92%). However, the high susceptibility of power Doppler to motion artefacts requires expert adjustments of the technical parameters, and, if anything, increases observer dependency. The resistive index within the intramural vessels of the gallbladder which can now be measured by ultrasound techniques does not differentiate between inflamed and non-inflamed gallbladders.

Gallbladder scintiscanning using iminodiacetic acid derivatives (HIDA, PIPIDA scans) can be used to confirm a non-functioning gallbladder and is regarded as the most accurate test of acute cholecystitis with a sensitivity of 97% and a specificity of 87%. A normal gallbladder scintiscan is virtually 100% accurate in excluding acute cholecystitis. The presence of pericholecystic uptake of the isotope is a valuable secondary sign in the diagnosis of acute cholecystitis and correlates with the presence of gangrenous cholecystitis or gallbladder perforation.

Intravenous cholangiography has been superseded by sonography in the diagnosis of acute cholecystitis. CT scanning is useful in complicated cases but ill advised in the majority in view of the radiation dosage. Good diagnostic accuracy for acute cholecystitis has been reported with MRI.

Acute acalculous cholecystitis

Acute acalculous cholecystitis accounts for up to 5–14% of cases of acute cholecystitis. It is seen most commonly in critically ill patients. The acute inflammation of the gallbladder arises in the absence of gallstones, although biliary sludge is often present. Although most commonly encountered in critically ill elderly patients, acute acalculous cholecystitis has also been reported in children. The disorder usually occurs during the course of a prolonged serious illness, e.g. multiple trauma, following major surgical intervention, in patients with extensive burns, severe sepsis and drug overdosage. The risk factors that predispose to the development of acute acalculous cholecystitis are:

- blood volume depletion
- prolonged ileus
- morphine administration exceeding 6 days
- intravenous hyperalimentation
- multiple blood transfusions
- sepsis
- starvation.

However, acute acalculous cholecystitis does not always arise on a background of a critical illness, and in one report more than 70% of patients suffering from the condition presented *de novo*. Most of these patients were elderly men with atheromatous vascular disease and 15% were diabetic.

Pathology

The exact pathology is not known but the available evidence suggests that the inflammation develops as a consequence of prolonged distension of the gallbladder, bile stasis and inspissation (biliary sludge), which results in mucosal injury and thrombosis of vessels of the seromuscular layer of the gallbladder. These critically ill patients are often medicated with narcotics, placed on ventilators and receive hyperalimentation that contribute to biliary stasis and functional obstruction of the cystic duct obstruction. The thrombosis is thought to be initiated by the activation of factor XII. Some have suggested that the development of acute acalculous cholecystitis may be related to a hypersensitivity reaction to concomitant antibiotic therapy because of the frequent presence of substantial eosinophilic infiltration of the mucosa of the gallbladder. Culture of the aspirated gallbladder bile from these patients is positive in only 38% of cases. Thus the inflammation is predominantly chemical in origin. In the fully developed condition, the gallbladder shows marked oedema of the seromuscular layer, mucosal ulceration, sloughing and focal necrotic areas. Gangrene of the gallbladder develops in approximately 40–60% of patients who have an associated increased risk for perforation. Mortality ranges from 6% to 44% but can be reduced by early diagnosis and therapy.

Clinical features

The diagnosis of acute acalculous cholecystitis is often difficult, especially in critically ill patients receiving narcotics and on artificial ventilation. The early manifestations include fever, leucocytosis and tenderness in the right hypochondrium. However, the diagnosis of acalculous cholecystitis is difficult to make clinically and by ultrasound, because gallstones are absent and the sonographic Murphy's sign may not be detected because of diminished mental status, medication and comorbid illness. In the series reported by Cornwall and colleagues, only 50% of patients who had acalculous cholecystitis had a positive Murphy's sign. The ultrasound diagnosis is, therefore, based on distension of the gall bladder in a suspicious clinical setting, the presence of intraluminal debris, gallbladder tenderness when present (~50%) and gallbladder wall thickness greater than 4 mm, pericholecystic fluid or subserosal oedema without ascites, intramural gas or sloughed mucosal membrane. However, ultrasonography and isotope scintiscanning are less accurate for this condition than their diagnostic yield for acute calculous cholecystitis. CT provides better diagnostic specificity in patients at risk of the condition by confirming the presence of pericholecystic inflammation in patients with a thick gallbladder wall demonstrated by ultrasonography.

Treatment

Emergency surgical therapy is essential for survival. In the absence of significant gangrene, cholecystostomy (performed percutaneously, laparoscopically or by mini-cholecystostomy) is increasingly favoured, particularly in critically ill patients. Follow-up of patients treated with cholecystostomy has confirmed return to normal gallbladder function in the majority of these patients. Established gangrene requires total or subtotal cholecystectomy. Some advocate change of antibiotic regimen in view of the unproven hypersensitivity theory.

Acute emphysematous cholecystitis

This fulminant form of acute cholecystitis is fortunately rare, accounting for less than 1% of all complicated cases of acute cholecystitis. It is caused by a mixed polymicrobial infection which includes gas-forming bacteria (*E. coli, Clostridium welchii*, aerobic and anaerobic streptococci). The gallbladder may or may not contain stones. Thrombosis of the cystic artery has been implicated in the development of acute emphysematous cholecystitis.

Clinical features

Acute emphysematous cholecystitis occurs predominantly in males (70%) and has a special predilection for diabetic individuals up to 40% in reported series. The affected patients do not often have gallstones. The clinical picture is that of severe rapidly oncoming upper abdominal emergency with prostration fever and marked toxicity. The clinical course is rapidly progressive, with a 75% incidence of gallbladder gangrene and a 20% incidence of gallbladder perforation by the time of diagnosis. The presence of air within the gallbladder lumen, its wall or the biliary tree on the plain radiograph is diagnostic. Emphysematous cholecystitis can be recognized on ultrasound examination by the extremely echogenic gas which casts a distal shadow and layers non-dependently within the gallbladder lumen. If the findings of the ultrasound or plain abdominal film are equivocal, CT is used to confirm the diagnosis.

Treatment

The antibiotic regimen of choice in these patients is a combination of penicillin and aminoglycoside. Emergency surgical intervention is imperative and consists of cholecystectomy, which is usually carried out by the laparoscopic approach but in the knowledge that there is approximately a 30% conversion rate to an open procedure in patients with advanced complicated disease: gangrene, perforation and established peritonitis. The reported mortality of emphysematous cholecystitis is 15%.

Complications of acute cholecystitis

The important complications of all forms of acute cholecystitis are empyema, perforation and gangrene. All require urgent surgical intervention.

Empyema (suppurative cholecystitis)

Empyema of the gallbladder is an uncommon complication and has a reported incidence of 2–3% of all patients with gallstone disease. It presents as a tender mass in the right hypochondrium and usually affects elderly patients in whom systemic signs, including pyrexia and leucocytosis, may be minimal. Cultures of the gallbladder contents are positive in 80%. Empyema of the gallbladder doubles the mortality figures of cholecystectomy.

Gangrene

Patchy gangrene of the fundus of the gallbladder is encountered in 5–7% of patients with obstructive cholecystitis. It is more commonly encountered in elderly patients, diabetics and in patients with empyema of the gallbladder, acute acalculous cholecystitis and, especially, emphysematous cholecystitis. It may lead to localized or free perforation of the gallbladder. Gangrenous cholecystitis is defined histologically as coagulative necrosis of the mucosa or the entire gallbladder wall associated with acute or chronic inflammation. It occurs in up to 20% of patients who have acute cholecystitis and has an increased risk for perforation. Unfortunately ultrasound is non-specific for the diagnosis of gangrenous cholecystitis. This is because the sonographic Murphy's sign is absent in up to 75% of patients. A specific finding is the presence of intraluminal membranes or stranding caused by sloughing of the gallbladder mucosa, necrosis of the gallbladder wall or fibrinous exudates. This finding is present on ultrasound examination, however, in only 5% of patients.

Perforation

Perforation of the gallbladder occurs in 5–10% of patients who have acute cholecystitis, most often in association with gangrenous cholecystitis. The fundus is the most common site for perforation, because it has the least blood supply. Acute gallbladder perforation with an intraperitoneal bile leak will result in peritonitis but is much less common than subacute perforation, which typically leads to pericholecystic abscess formation. These abscesses may occur within or adjacent to the gallbladder wall in the gallbladder fossa, within the liver, parenchyma, or along the free margin of the gallbladder within the peritoneal cavity. A localized perforation may involve the duodenum with the development of a cholecystoduodenal fistula and resolution of the inflammatory episode. However, this bilioenteric fistula persists and passage of a large stone through this fistula may eventually cause gallstone ileus.

Free perforation resulting in generalized infected biliary peritonitis carries a high mortality, variously reported as 30–50%.

Treatment of acute cholecystitis

Initial management

This consists of intravenous fluid and electrolyte replacement, nasogastric suction, systemic antibiotics and parenteral analgesia. The patient is kept fasted to reduce the CCK release from the upper small bowel in order to minimize gallbladder stimulation. Although the inflammation is initially chemical, most surgeons will choose to use systemic antibiotics because of the risk of progression to an empyema and septic complications. Also, if surgery is performed, antibiotic prophylaxis will reduce the wound infection rate although this has recently been questioned in patients undergoing LC. As the organisms cultured from gallbladder bile are predominantly Gram-positive aerobes (*E. coli, Klebsiella* spp., *Streptococcus* spp.), a third-generation cephalosporin is the antibiotic of choice. Anaerobes such as *B. fragilis* and *Clostridium perfringens* are associated with more severe, mixed infections particularly in elderly people. These require combination chemotherapy using metronidazole with an aminoglycoside and/or penicillin. The diagnosis of acute cholecystitis should be confirmed during this initial 12–24 hour period of stabilization by ultrasonography or gallbladder scintiscanning.

Opinions still differ regarding the definitive treatment of acute cholecystitis. The management depends on whether the inflammatory condition is progressive and life-threatening or the cholecystitis is mild and resolving.

Severe progressive disease

The timing of surgery is dictated by the severity of the attack. Box 25.1 summarizes the indications for emergency or urgent surgical intervention.

In these patients surgical intervention is carried out under antibiotic cover active against both Gram-negative aerobes and anaerobes (cephalosporin + metronidazole or piperacillin, etc.). Traditionally, such patients have been managed by laparotomy using a midline epigastric incision and the open approach is still favoured by many in critically ill elderly patients. Hitherto, less than 10% of cases fell into this severe category but more recent reports from North America and Europe have highlighted an increasing proportion of acute severe cases requiring urgent surgical intervention (from 6–10% up to 25%). This has been attributed to a decrease in the number of patients undergoing elective cholecystectomy though this has been reversed with the advent of LC, an increasingly aged population, and a rise in the actual incidence of acute complications of gallstone disease.

The exact procedure depends on the operative findings. In patients with a tense empyema, preliminary decompression of the gallbladder contents using a Mayo–Ochsner suction trocar–cannula inserted through a purse-string suture in the fundus should precede the cholecystectomy which, in the acute situation, is best performed by the retrograde technique (starting at the fundus). This allows easier identification of the cystic duct and, thereby, reduces the risk of bile duct damage. At times, the precarious condition of the patient precludes a lengthy operation or the anatomy may be so obscured by the inflammatory mass as to render the cholecystectomy hazardous. In these situations, a cholecystostomy should be performed. The gallbladder contents are evacuated, any gangrenous patches of its walls are excised and a 22–24 Fr Malecot catheter is inserted into the organ, which is closed round it by a purse-string suture. The catheter is then brought out through a separate stab wound. In these patients a cholecystectomy is advisable at a later stage unless the patient is elderly or has severe comorbid cardiorespiratory disease, because of the risk of recurrence of gallstones and symptoms. Moreover, the incidence of carcinoma of the gallbladder in patients who had previously undergone cholecystostomy is appreciable (7%).

Subtotal cholecystectomy is performed as an alternative approach to cholecystostomy in patients in whom formal cholecystectomy is considered hazardous. In this procedure, the posterior wall of the gallbladder is left *in situ,* attached to the liver bed, and the cystic duct is secured from within the gallbladder lumen by a purse-string suture (Figure 25.45).

In all instances specimens of bile and pus are obtained for bacteriological culture. Pus is thoroughly evacuated and peritoneal lavage, preferably with an antibiotic solution, carried out when gross peritoneal sepsis is found. Adequate drainage of the gallbladder bed is still considered advisable but is no substitute for thorough peritoneal toilet. All these patients are at risk from Gram-negative bacteraemia. They require a full course of antibiotic therapy for a minimum of 7 days. The results of the culture of operative specimens of bile and pus may dictate changes in the antibiotic regimen.

BOX 25.1 Indications for emergency surgical intervention in patients with acute cholecystitis

- Progression of the disease despite conservative treatment
- Failure to improve within 24 hours especially in patients >60 years
- Presence of an inflammatory mass in the right hypochondrium
- Detection of gas in the gallbladder/biliary tract
- Established generalized peritonitis
- Development of intestinal obstruction

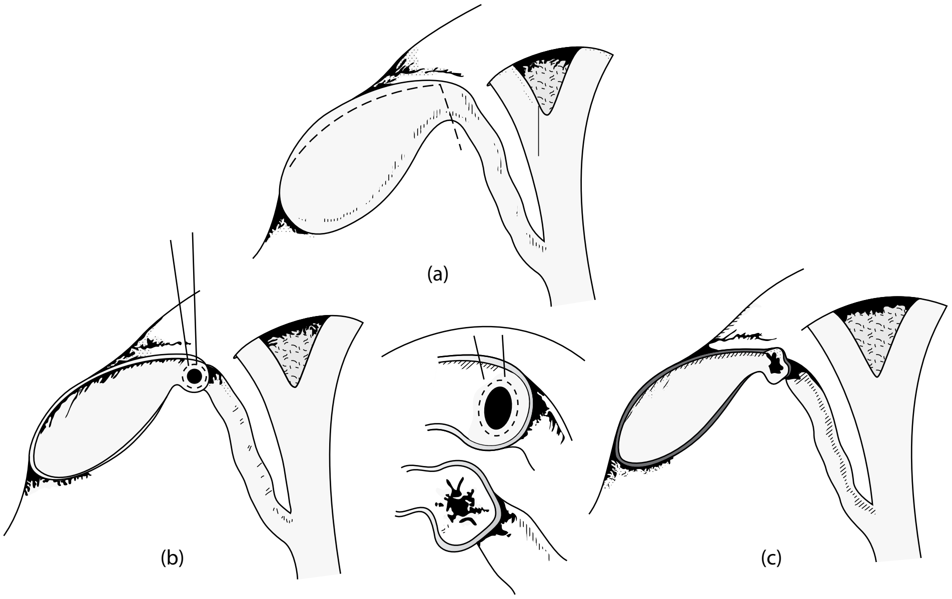


Figure 25.45 Technique of subtotal cholecystectomy described by Bormann and Terblanche: (a) the broken line shows the extent of the cholecystectomy – the posterior wall of the gallbladder is left attached to the liver; (b) a purse-string suture is inserted around the cystic duct orifice; (c) the purse-string suture has been tied with closure of the cystic duct orifice.

Laparoscopic management of severe acute disease

Although the open surgical management outlined above is still considered by many to be the standard surgical management of these severely ill patients, there is a strong case, and indeed an increasing trend, for laparoscopic management. This is intended in the first instance as a diagnostic inspection to assess the severity of the disease. In the presence of established gangrene and perforation, open surgical management as outlined in the preceding section is indicated. Otherwise, a laparoscopic insertion of a self-retaining catheter will effectively drain the inflamed gallbladder and tide the patient over a critical illness. Once this has resolved and if the patient's condition permits, a cholecystectomy (laparoscopic or otherwise) is performed. If the patient's cardiorespiratory status precludes an elective surgical intervention, then gallstone extraction, fragmentation or MTBE dissolution through the established cholecystostomy tract is performed.

Mini-cholecystostomy of Burhenne and Stoller

This is an equally valid approach in poor-risk patients with severe acute disease. After the position of the fundus of the inflamed gallbladder is located by ultrasound, a small incision is made over it. Stay sutures are inserted. The gallbladder contents are aspirated and sent for culture and a Foley catheter is inserted into the gallbladder lumen and held in place by a purse-string suture. This approach, which can be performed under local anaesthesia and sedation, is quick and extremely safe.

Mortality of severe acute cholecystitis

The overall reported mortality of acute cholecystitis is 3%. The mortality in the elderly is higher (10%) and more than half of the deaths in patients over 65 years are secondary to cardiovascular and respiratory complications.

Established non-progressive disease

These form the majority and the acute obstructive cholecystitis usually resolves with conservative treatment. There are two management options:

- delayed (interval, subsequent admission) cholecystectomy
- early (same admission) cholecystectomy.

The interval approach is the traditional one and entails conservative management of the acute episode with discharge of the patient after complete resolution of the attack. Subsequently, the patient is admitted some 2–3 months later for an elective cholecystectomy. The rationale for this treatment is that in most instances the raised pressure within the gallbladder lumen lifts the walls of the organ off the impacted stone which then dislodges and falls into the lumen with resolution of the inflammation, the view being held that it is safer to operate several weeks after the acute inflammatory episode has subsided.

Early cholecystectomy is increasingly favoured in the management of acute cholecystitis. It must be distinguished from emergency cholecystectomy. Following initial conservative management and confirmation of the diagnosis as outlined previously, the patient is operated electively (scheduled urgent) on the next available operating list or within a few days of admission. The results of several prospective clinical trials comparing early vs interval cholecystectomy have shown clear benefits from early cholecystectomy performed during the same hospital admission in fit patients (ASA I and II). These include less time spent in hospital and lower cost of treatment. Fears that an early cholecystectomy is a more hazardous procedure have proved groundless, in particular the incidence of complications including missed common duct stones and mortality rates reported in these prospective trials have been similar. On the other hand, the delayed management has several disadvantages which include:

- failure of conservative treatment in 13%
- premature readmission with a further attack while waiting for elective cholecystectomy (13%)
- patient defaulting after discharge (10%).

More recently, a randomized trial of early vs interval LC for acute cholecystitis has demonstrated similar results. In particular there was no significant difference in the conversion rates (early 21% vs interval 24%), similar morbidity but a significantly ($p < 0.001$) shorter hospital stay, 7.6 vs 11.6 days, although the operating time was longer for the early group, i.e. 122 vs 106 minutes. Similar benefits have been reported by another randomized trial comparing early vs interval LC for acute cholecystitis. Thus early LC carries both medical and socioeconomic benefits over interval LC for acute cholecystitis.

An aggressive policy of early cholecystectomy is indicated in elderly and diabetic patients unless they have comorbid significant heart disease as these patients often have gangrenous disease.

Laparoscopic vs open cholecystectomy

As with the elective situation, LC offers significant advantages over open cholecystectomy for acute cholecystitis, and early LC has rapidly become the treatment of choice for this condition. Aside from several retrospective reports indicative of reduced morbidity and hospital stay, the benefit of LC vs open cholecystectomy for acute cholecystitis has been confirmed by a prospective randomized clinical trial. This study showed a significant reduction in the postoperative morbidity in the LC arm (only 3% minor complications vs 23% major complications and 19% minor complications in the open cholecystectomy group). In adopting LC as the routine option, it must be stressed that the need for conversion is encountered in 5–15% of all cases and is higher in LC for acute cholecystitis. An early decision should be made to convert electively in the presence of obscured anatomy. This is far better than persistence with a difficult operation with enforced conversion because of the onset of an intraoperative complication. These patients are at risk of severe postoperative complications.

Thus the valid approach is a flexible one. The procedure starts with an exploratory laparoscopy to assess technical difficulty of the operation with particular reference to the structures in the triangle of Calot. A large distended gallbladder should be aspirated and lifted by a retractor rather than grasped. Large stones impacted in Hartmann's pouch that cannot be dislodged may prove problematic. The practical axiom is a simple one, i.e. if adequate exposure for a safe dissection cannot be obtained, the case should be converted. In some cases, a fundus first dissection of the gallbladder may be required.

The case for routine fluorocholangiography is much stronger when cholecystectomy is performed for acute cholecystitis as these patients are much more commonly jaundiced (the inflammatory oedema or by concomitant ductal calculi). The cholangiogram should outline the entire biliary tract (intra- and extrahepatic). It ensures safe occlusion of the cystic duct stump without compromise of the common hepatic or common duct and differentiates stones from distortion caused by the inflammatory oedema.

Mucocele of the gallbladder

Mucocele of the gallbladder is usually encountered in elderly patients and presents with a painless mass in the right hypochondrium. Reported data indicate that about 3% of all pathological gallbladders in adults are mucoceles, although the true prevalence may be higher because of the varying criteria used to define the condition. Most commonly mucoceles are associated with solitary gallstones impacted in Hartmann's pouch.

Pathology

The usual underlying pathology is due to longstanding obstruction to the gallbladder which becomes distended with mucus such that the organ may enlarge considerably. The obstruction to the outflow from the gallbladder results in the slow reabsorption of bile and bile pigment but the gallbladder mucosa continues to secrete clear and watery or mucoid fluid (white bile) in copious amounts. Although the gallbladder wall may be of normal thickness, in longstanding cases there is atrophy of the mucosa such that the wall becomes thin and semitransparent. However, wall thickening can develop with recurrent attacks of infection (cholecystitis). The contents of mucocele of the gallbladder are usually sterile, although infection may supervene leading to empyema of the gallbladder. Progressive overdistension may result in gangrene and/or perforation of the gallbladder, with ensuing pericholecystic collection or peritonitis. Other causes of mucocele include:

- resolved acute cholecystitis
- polyps or malignancy of the gallbladder
- compression of the neck or cystic duct by lymph nodes or inflammatory fibrosis
- congenital narrowing of the cystic duct
- obstruction by parasites, e.g. *A. lumbricoides*.

Clinical features

There may or may not be a history of acute pain indicative of biliary colic or mild acute cholecystitis. The clinical features of gallbladder mucocele include right upper quadrant or epigastric pain and discomfort, nausea and vomiting. Fever and rigors indicate the onset of infection with the development of empyema of the gallbladder. Jaundice is unusual except in patients with coexisting bile duct obstruction either by stones or by extrinsic compression (Mirizzi syndrome). On palpation, the distended gallbladder is usually palpable and tender. In uncomplicated cases the laboratory test results are normal or just within the upper limit of the normal range for the laboratory. Plain radiograph of the abdomen shows a soft-tissue density, globular shadow in the subhepatic region and ultrasonography is usually diagnostic: thin distended gallbladder with minimal wall thickening, and impacted stone in the neck/Hartmann's pouch.

Treatment

The treatment is cholecystectomy which can be performed safely by the laparoscopic approach. It is important that the thin-walled distended gallbladder is decompressed by aspiration before it is grasped by any laparoscopic forceps to prevent rupture and escape of contents into the peritoneal cavity.

Acalculous chronic gallbladder disease: hyperplastic cholecystoses

Chronic inflammation of the gallbladder in the absence of gallstones is due to adenomyomatosis or cholesterolosis of the gallbladder and these may be conveniently grouped as acalculous chronic gallbladder disease, sometimes also referred to as hyperplastic cholecystoses. Both conditions are common and involve various layers of the gallbladder wall, but are usually asymptomatic. However, they can give rise to vague symptoms not dissimilar to those of chronic cholecystitis. Oral cholecystography shows no abnormality of the gallbladder in 50% of these patients, or is reported as demonstrating poor function or unusual appearances. Thickening and 'polypoid' lesions/diverticula may be documented by high-resolution ultrasound. In many instances, the diagnosis is made on pathological examination of the excised gallbladder. In general, the results of cholecystectomy for these conditions have been difficult to evaluate but, in the presence of persisting symptoms, most surgeons would advise operation.

Adenomyomatosis of the gallbladder

Adenomyomatosis (diffuse, focal or polypoid) is found in 8.7% of surgical gallbladder specimens. This is variously named as adenomatous hyperplasia, diverticulosis of the gallbladder and cholecystitis glandularis proliferans. It is thought to represent a developmental defect that results in hyperplasia of the smooth muscle bundles with sacculation or diverticulum formation of the epithelial lining (Rokitansky–Aschoff sinuses). The cholecystogram (rarely performed nowadays) may be normal or the late film taken after a fatty meal may demonstrate either the mural diverticula (Figure 25.46) or a concentric narrowing of the fundus.

Adenomyomatosis involves the mucosa and the muscular and connective tissue layers of the gallbladder wall. The epithelium and muscular layers proliferate, and the subsequent invagination of the mucosa into the gallbladder wall produces intramural diverticula (med Rokitansky–Aschoff). These diverticula may accumulate bile, cholesterol crystals, or even stones. On ultrasound examination they may be anechoic if large enough and contain bile but more frequently the diveticula are small and contain cholesterol, biliary sludge or small gallstones that create echogenic foci, often with ring-down or comet tail reverberation artefacts.

There are three distinct forms of adenomyomatosis. The most common form of adenomyomatosis is a focal polypoid lesion, also known as an adenomyoma, typically located at the tip of the gallbladder fundus. The segmental form consists of localized gallbladder wall thickening that typically narrows the gallbladder body in an hourglass configuration. Diffuse adenomyomatosis involving the entire gallbladder wall is less common than focal or segmental diseases.

Treatment

This is needed only in symptomatic patients in whom other pathologies have been excluded by appropriate investigations, and consists of LC, which may or may not relieve their symptoms.

Cholesterolosis of the gallbladder

In this condition, the epithelial cells and macrophages within the gallbladder mucosa become laden with cholesterol and lead to the formation of numerous lipid deposits. Chronic inflammation of the adjacent mucosa then leads to a striking appearance of the interior of the gallbladder, which has been aptly described as the 'strawberry gallbladder'.

Cholesterolosis, which may be diffuse or polypoid, has been reported in up to 25% of surgical specimens. The condition is caused by deposition of lipid-laden macrophages in the lamina propria of the normal mucosal epithelium of the gallbladder wall. The diffuse form, which is more common, is more likely to cause symptoms. Cholesterol polyps account for 20% of cholesterolosis but constitute approximately one-half of all gallbladder polyps. They are usually less than 1.0 cm in size, often multiple, and have no malignant potential. On ultrasound they appear as brightly echogenic, round or lobulated, immobile, non-shadowing masses abutting the gallbladder wall. The condition is very rarely symptomatic.

Ductal calculi and cholangitis

The majority of ductal calculi are found in the common bile duct and, in the West, only an estimated 5% of ductal calculi are located

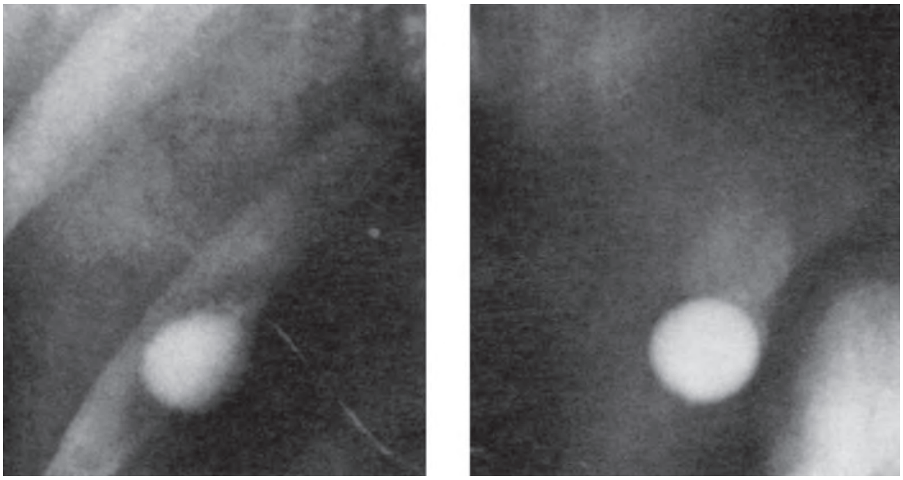


Figure 25.46 Oral cholecystogram: a concentric narrowing of the fundus is the striking feature but on closer inspection, intramural diverticula are discernible. When encountered, these are pathognomonic of adenomyomatosis of the gallbladder. Often, however, the oral cholecystogram is normal.

in the intrahepatic ducts (more commonly the left), although multiple intrahepatic calculi are common in Eastern countries, especially where parasitic infestations and recurrent pyogenic cholangitis are endemic. Ductal calculi may arise as follows:

- as secondary calculi from migration of gallstones
- as primary calculi arising *de novo* within the bile ducts.

Primary ductal calculi are brown pigment stones. Their composition and infective aetiology are discussed earlier in this chapter. In surgical practice, brown pigment stones are often encountered in the following conditions:

- stasis in the biliary tract caused by strictures, tumours and sclerosing cholangitis
- parasitic infestations
- recurrent pyogenic cholangitis (Asia)
- indwelling endoprostheses/internalized clips and non-absorbable sutures.

Less commonly, pigment stones form around metal clips and non-absorbable sutures when these become incorporated inside the bile duct. The important predisposing condition to infection and brown pigment stone formation is stasis and this requires correction in addition to removal of the calculi and eradication of any infection.

Clinical manifestations of ductal calculi

Although 15–20% of patients with stones in the common bile duct are asymptomatic, the majority present sooner or later with severe symptoms, and by and large incur a significant morbidity as the pathological potential of ductal calculi is high and may contribute to the death of the patient. Ductal calculi may present with:

- recurrent bouts of biliary colic accompanied by intermittent jaundice
- episodic upper abdominal pain and dyspepsia (Figure 25.47)
- stone impaction with progressive jaundice
- cholangitis
- gallstone pancreatitis
- secondary biliary cirrhosis and portal hypertension.

Management of ductal calculi

Role of endoscopic sphincterotomy and stone extraction

The current orthodox treatment for ductal calculi is by endoscopic sphincterotomy and stone extraction; in patients requiring cholecystectomy for symptomatic gallstone disease, endoscopic stone extraction is performed before the operation preferably during the same hospital admission. However, the increasing efficacy and usage of laparoscopic stone clearance for patients harbouring ductal calculi is challenging the role of preoperative endoscopic stone extraction in centres with the necessary expertise. In addition, there is increasing concern regarding sphincterotomy in patients below the age of 50 years. The short-term morbidity of endoscopic stone extraction ranges from 5% to 10%. It has a median failure rate of 10% and a mortality of 0.5–1.0%. In addition, endoscopic sphincterotomy carries a long-term morbidity that includes recurrence of common duct stones (average of 11%), and stenosis (3% in the absence of papillitis and 30% if papillitis is present). Further, the sphincterotomy results in bacterial colonization of the extrahepatic biliary tract.

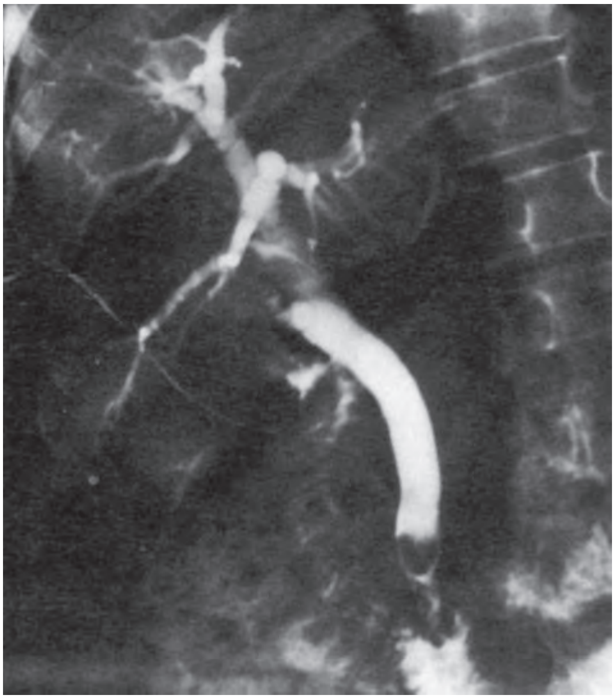


Figure 25.47 Primary ductal calculi situated at the lower end of the bile duct 13 years after a cholecystectomy. The patient complained of dyspepsia and episodic abdominal pain. Although the patient was not jaundiced, the liver function tests showed a mild elevation of the serum bilirubin and raised alkaline phosphatase.

There is agreement that the two-stage approach is indicated in poor-risk patients including those with established cholangitis or severe pancreatitis, but these constitute only 30% of patients with ductal calculi. In fit patients with ductal stones, single-stage laparoscopic surgical treatment (laparoscopic stone extraction and cholecystectomy) is gaining favour among laparoscopic surgeons, especially as the evidence from four randomized clinical trials is that the two-stage sequential management of patients with ductal stones is inferior (increased overall morbidity) to single-stage surgical treatment. A large clinical trial involving 300 randomized patients showed the same efficacy, morbidity and mortality but a significant reduction in the hospital stay in the single-stage laparoscopic arm. Undoubtedly, the single-stage laparoscopic approach requires a lesser resource and should replace the two-stage approach on the grounds of equal efficacy and safety but diminished costs. On average in 50% of patients undergoing ERCP for suspected ductal calculi, the ERCP is normal. The persistence of the two-stage approach is attributed to reluctance of gastroenterologists to relinquish this treatment and the limited number of centres undertaking laparoscopic ductal stone extraction routinely – regrettably the two are interlinked. If evidence-based treatment is to be followed, then endoscopic stone extraction should reserved for:

- poor-risk patients
- patients with cholangitis
- patients with severe pancreatitis

- some patients with failed laparoscopic stone extraction as an alternative to conversion
- retained or recurrent stones after cholecystectomy.

In elderly or poor-risk patients, endoscopic sphincterotomy alone leaving the gallbladder *in situ* has been recommended in the past. However, subsequent clinical trials comparing this approach with definitive single-stage treatment (open or laparoscopic) have shown similar morbidity and mortality rates and a high incidence (30%) in patients with *in situ* gallbladder of biliary symptoms/cholecystitis requiting subsequent cholecystectomy. Thus endoscopic sphincterotomy alone is no longer favoured unless the patient is totally unfit for surgery (Figure 25.48).

Laparoscopic extraction of ductal calculi

The techniques of laparoscopic ductal stone clearance include:

- transcystic duct extraction
- direct supraduodenal common bile duct exploration
- rendezvous laparoendoscopic approach.

Transcystic duct extraction

This is the least invasive technique and has the undoubted merit of leaving the extrahepatic biliary tract including the choledochal sphincter intact. Thus its morbidity is low and the average postoperative hospital stay is short (3 days). It is indicated for small floating stones up to 5 mm diameter, where its reported efficacy in achieving ductal stone clearance ranges from 80% to 95%. There are two commonly used techniques:

- the choledochoscopic visually guided method
- the radiologically guided wire basket trawling technique.

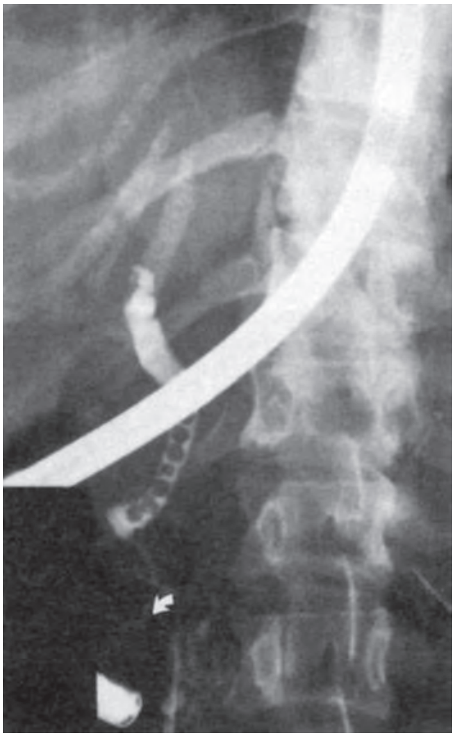


Figure 25.48 Endoscopic papillotomy for multiple ductal calculi in an elderly patient. The wire of the sphincterotome is visible above the tip of the endoscope.

The choledochoscopic technique requires dilatation of the cystic duct before insertion of the mini-choledochoscope through the cystic duct into the common bile duct. The stones are visualized and trapped using a wire basket introduced through the instrument channel of the endoscope and extracted through the cystic duct. Some advocate delivering the trapped stones through the sphincter of Oddi into the duodenum where they are released. This is, however, not generally recommended because of the risk of pancreatitis. A completion cystic duct cholangiogram is performed, and if this is normal the cystic duct is ligated and the gallbladder removed. If there is any doubt about residual fragments, a cystic duct drainage cannula (Cook) is inserted into the common bile duct and tied to the cystic duct with two Roeder catgut extracorporeal slip knots. A postoperative cholangiogram is performed 24 hours later. If this is normal, the cystic duct drainage cannula is capped with a Luer lock and covered with an occlusive dressing before the patient is discharged usually on the third postoperative day. The cannula is removed as an outpatient 10–14 days later.

The radiologically guided trawling technique is quicker and simpler than the above. Furthermore, it has the added advantage of obviating the need for dilatation of the cystic duct. It consists of the insertion of a soft pleated four-wire basket in the closed position through the cystic duct into the lower end of the common bile duct. Once in position with the tip among the stones, the basket is fully opened and then withdrawn slowly. No attempt is made to close the basket as the wires of this are folded over the stones by the narrow cystic duct during the trawling process (Figure 25.49a–d). With experience, the radiological exposure is minimal. The trawling process is repeated until complete stone clearance is achieved.

Laparoendoscopic approach

This combined approach entails insertion of a guide wire by the laparoscopic surgeon through the cystic duct down the common bile duct into the duodenum. The endoscopist then performs a sphincterotomy and stone extraction with the patient in the supine position. The reported experience with the laparoendoscopic technique although limited has been entirely favourable, but the technique requires additional resource and the availability of an experienced flexible endoscopist. The perceived risk of acute postoperative pancreatitis with this technique has not materialized in any of the reported series.

Direct supraduodenal common bile duct exploration

This is indicated for large or occluding stones. The technique is similar to open exploration of the common bile duct. The dissection is minimal and consists of exposure of the anterior wall of the common bile duct after downward displacement of the duodenum. No stay sutures are inserted. A choledochotomy is made in the supraduodenal segment. The size of the choledochotomy should be approximately half to three-quarters the maximum diameter of the largest stone. Avoidance of a large choledochotomy reduces the amount of intracorporeal suturing needed, and, in view of the elasticity of the common bile duct, the opening can be stretched to accommodate large stones. The extraction manoeuvres include milking the duct from below upwards with an instrument on side, balloon dislodgement and direct visual wire basket extraction after the insertion of a 4 mm flexible choledochoscope. The stones delivered through the choledochotomy are removed by the Semm's spoon forceps. Following ductal clearance, biliary drainage is advisable. This can be performed by the orthodox T-tube (Figure 25.50), or preferably through a cystic duct drainage cannula (Figure 25.51). This is inserted until the S-shaped perforated terminal segment is well inside the common bile duct, after which it is secured to the cystic duct by two Roeder catgut sutures. Primary closure of the choledochotomy is carried out, during which the common bile duct is constantly irrigated with saline via the cystic duct drainage cannula.

In either case, a completion cholangiogram is essential to ensure ductal clearance and a subhepatic drain is necessary. The cystic duct drainage cannula or the T-tube is firmly anchored to the abdominal wall and connected to a biliary drainage bag. When a cystic duct drainage cannula is used, the postoperative cholangiogram is carried out on the second day; if this is normal, the cannula is sealed with a Luer lock and covered with an occlusive dressing. In the absence of any bile leakage from the subhepatic drain over the next 24 hours, the drain is removed and the patient discharged the next day with the cystic duct drainage cannula *in situ*. This is removed 7 days later as an outpatient. If a T-tube has been inserted, the postoperative cholangiogram is carried out on the seventh postoperative day, and, if normal, the tube is clamped and provided there is no leakage from the subhepatic drain and the patient has no symptoms or fever over the next 24 hours the T-tube is removed. The subhepatic drain is removed a day later. In the elderly, diabetics and patients on immunosuppressive drugs, the T-tube is kept in place for at least 2 weeks as the maturation of the T-tube tract is impaired in these patients. Thus early removal may lead to biliary peritonitis.

Patients with ductal calculi discovered during elective cholecystectomy (unsuspected ductal calculi)

These account for 4–10% of cases and are discovered by routine IOFC. They are usually small floating calculi without dilatation of the common bile duct. As these patients have no obstruction and normal liver function tests, some surgeons have advocated no intervention other than the cholecystectomy on the grounds that some of these calculi will pass spontaneously and if and when

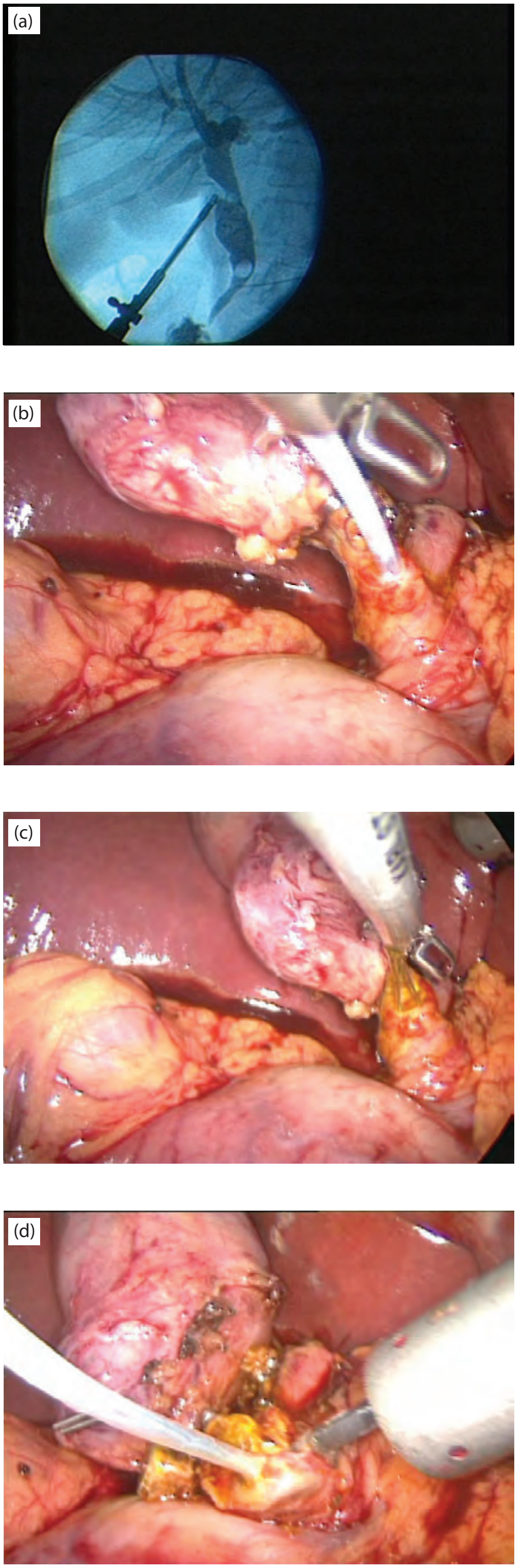


Figure 25.49 (a) Ductal calculi on the intraoperative cholangiography. (b,c) Transcystic duct clearance with a pleated soft wire basket.

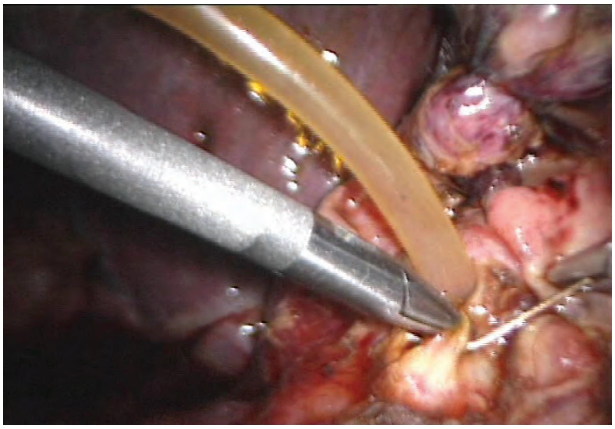


Figure 25.50 Laparoscopic common bile duct exploration with closure of the common bile duct around a T-tube.

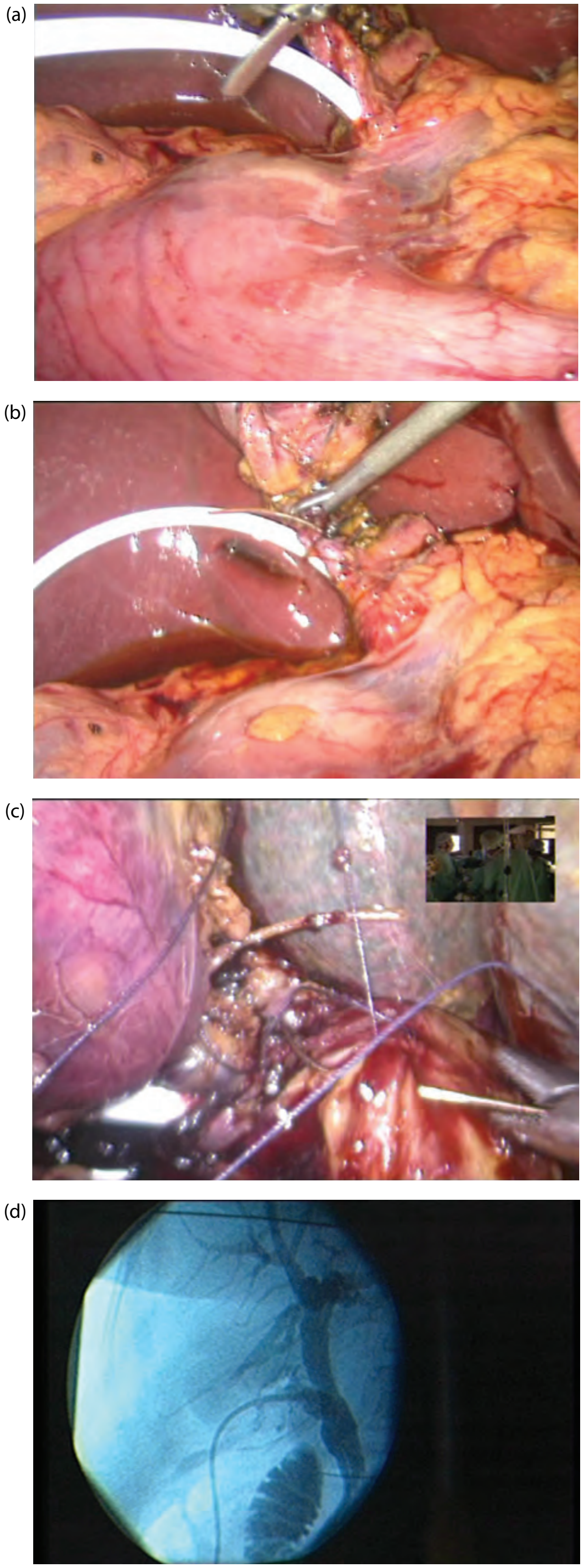


Figure 25.51 Decompression of the extrahepatic biliary tract by the Cuschieri cystic duct drainage cannula with primary closure of the choledochotomy: (a) a drainage cannula is inserted and fixed to the cystic duct stump; (b) primary closure of the choledochotomy; (c) completion cholangiogram through the cystic duct drainage cannula.

they become symptomatic or complications arise after surgery, endoscopic sphincterotomy can be performed. This policy is ill founded for several reasons. In the first instance, there are no hard data as to the frequency of spontaneous passage without complications. Furthermore, it is impossible to predict uncomplicated stone passage in the individual patient. This expectant policy will incur a definite morbidity from jaundice, cholangitis or pancreatitis and some (albeit few) patients will die as a result of one of these complications. The appropriate management of these patients, based on the known pathogenic potential of ductal calculi, is removal either at the time of LC (transcystic duct extraction) or by postoperative endoscopic sphincterotomy.

Ductal calculi discovered soon after cholecystectomy and exploration of the common bile duct

These are referred to as missed, retained or residual stones. The incidence varies from 2% to 15% and averages 8%. Routine completion choledochoscopy/cholangiography virtually abolishes this complication. Retained ductal calculi following biliary tract surgery are either diagnosed in the immediate postoperative period by the postoperative T-tube cholangiogram or present with recurrent symptoms usually within 2 years of cholecystectomy without exploration of the common bile duct. Ductal stones presenting beyond this interval are generally considered to be of the primary variety.

Certain general considerations apply with regard to the management of patients with residual calculi following biliary tract surgery. Urgent intervention is not indicated if the liver biochemistry is normal, the patient is asymptomatic, and the T-tube cholangiogram shows no organic disease or significant dilatation. Spontaneous passage is likely if the calculi are small (less than 3 mm) and may be aided by simple measures such as T-tube clamping. If the patient tolerates clamping and providing no untoward symptoms or complications develop, such a conservative approach can be continued for a few weeks, at the end of which time the situation is reviewed radiologically.

The various methods available for the non-surgical management of retained stones are:

- flushing
- dissolution
- percutaneous stone extraction via the T-tube tract
- endoscopic sphincterotomy and stone extraction.

The first three options are applicable only to patients with an indwelling T-tube whereas endoscopic stone extraction can be used in all patients. All the above methods are performed under antibiotic cover because of the risk of cholangitis and septicaemia.

Flushing is usually carried out with saline, heparinized saline or lidocaine–saline solution. The technique, which is simple and effective if the stones are small (<0.3 mm), is performed by infusing the solution through the T-tube under manometric control to ensure that the pressure does not exceed 30 cmH_2O, as this can lead to cholangiovenous reflux and septicaemia. The efficacy of this simple method of treatment, which does not require any special expertise, can be enhanced when it is accompanied by pharmacologically induced relaxation of the sphincter of Oddi.

Cholate infusion can dissolve cholesterol stones but its efficacy is low and it has been replaced by mono-octanoin, which acts more rapidly and achieves complete stone clearance in 40%. The most effective agent for the dissolution of ductal calculi is MTBE, which is capable of achieving gallstone dissolution within hours of instillation.

Percutaneous stone extraction via the T-tube tract was initially performed by the Burhenne technique using a Dormia basket introduced via a specially designed steerable catheter to capture and extract the stone under fluoroscopic control. It has largely been replaced by the flexible choledochoscopic technique which is successful in 90–95% of cases. A 4–6 week period of maturation of the T-tube tract is required before the procedure can be performed safely. A guide wire is introduced into the common bile duct and the T-tube removed. Thereafter, the T-tube tract is dilated to allow the introduction of the narrow flexible choledochoscope. The retained stones are removed by means of a Dormia basket under visual control.

Endoscopic sphincterotomy with stone extraction is the most effective method of dealing with the problem of retained stones and can be performed in patients with and without T-tubes. Surgical management of missed stones is reserved for those patients in whom the above methods have failed or complications have developed during or after attempted endoscopic or percutaneous stone extraction.

Recurrent ductal calculi

Ductal calculi presenting 2 years or more after an operation are generally regarded as primary. One study has identified suture material in 30% of cases. This finding stresses the importance of avoiding non-absorbable material during operations on the biliary tract. Internalization of metal clips used to secure the medial end of the cystic duct during LC is now a well-recognized complication of this procedure. The exact pathology remains unclear, but it seems likely that the clip is placed too close to the common bile duct resulting in localized pressure necrosis. The internalized clip becomes covered with calcium bilirubinate to form a brown pigment stone. The patients who develop this condition present between 6 and 12 months after the procedure with jaundice and/or cholangitis. The condition is easily diagnosed on the ERCP films as the stone has a characteristic 'cat's eye' appearance.

The management of patients with recurrent ductal calculi depends on their age and general condition. Endoscopic sphincterotomy and stone extraction is the first-line treatment and surgery (open or laparoscopic) reserved if this approach fails. During surgery, the stones are removed atraumatically by means of biliary balloon catheters, stone-grasping forceps or Dormia baskets as described previously. A completion check by means of a choledochoscopic inspection or cholangiography abolishes or reduces the incidence of residual stones. Temporary biliary drainage is advisable and provides a means of decompression.

In other situations, recurrent ductal calculi are often multiple and associated with gross dilatation of the bile duct and in some cases obvious distal ductal stenosis. This may be primary (papillary stenosis) or be secondary to trauma inflicted by metal bougies introduced through the sphincter region at the time of exploration of the common bile duct. In patients with multiple ductal calculi, grossly dilated bile duct (>2 cm) or papillary stenosis, a drainage operation is indicated: choledochoduodenostomy or transduodenal sphincteroplasty. Opinions are divided as to the relative merits of these two procedures. However, sphincteroplasty carries a significant risk of pancreatitis and involves a sizeable duodenotomy. A transection choledochoduodenostomy (Figure 25.52) is preferable to the side-to-side anastomosis as it provides dependent drainage and avoids the complication of the inspissated sump syndrome. Transection choledochoduodenostomy can be performed laparoscopically and is a relatively straightforward procedure as the bile duct is grossly dilated in these patients (Figure 25.53).

Intrahepatic calculi (hepatolithiasis)

Intrahepatic stones (hepatolithiasis) are prevalent in Southeast Asia but rare in Western countries. They are defined as stones located in the biliary tract proximal to the confluence of

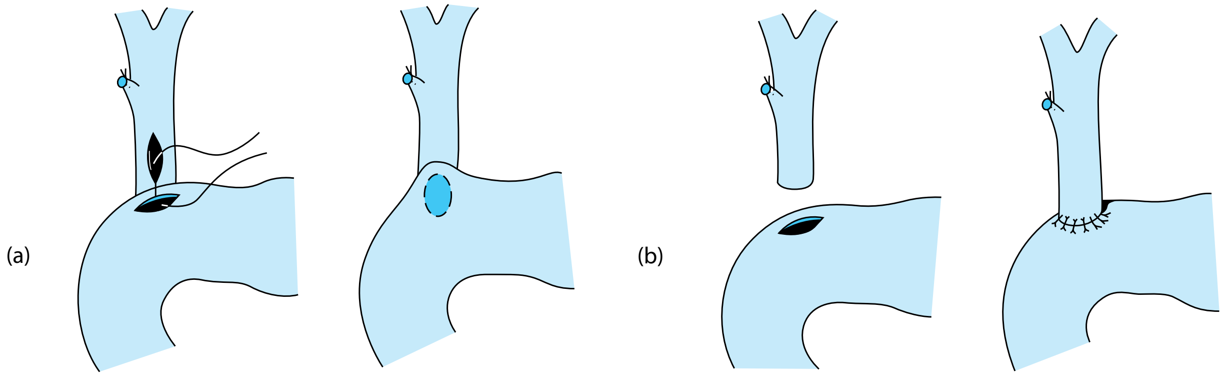


Figure 25.52 Choledochoduodenostomy: (a) lateral or side-to-side choledochoduodenostomy – this may result in the passage of food debris from the duodenum through the stoma into the lower end of the bile duct (distal to the anastomosis with the development of the inspissated sump syndrome which presents with cholangitis). (b) Diagrammatic technique of the procedure of transection choledochoduodenostomy. The bile duct is mobilized from the portal vein and hepatic artery. It is then transected as it enters the pancreas. The distal end is closed with a running suture and the proximal end is anastomosed in an end-to-side fashion to the duodenum at the junction of the first with the second part.

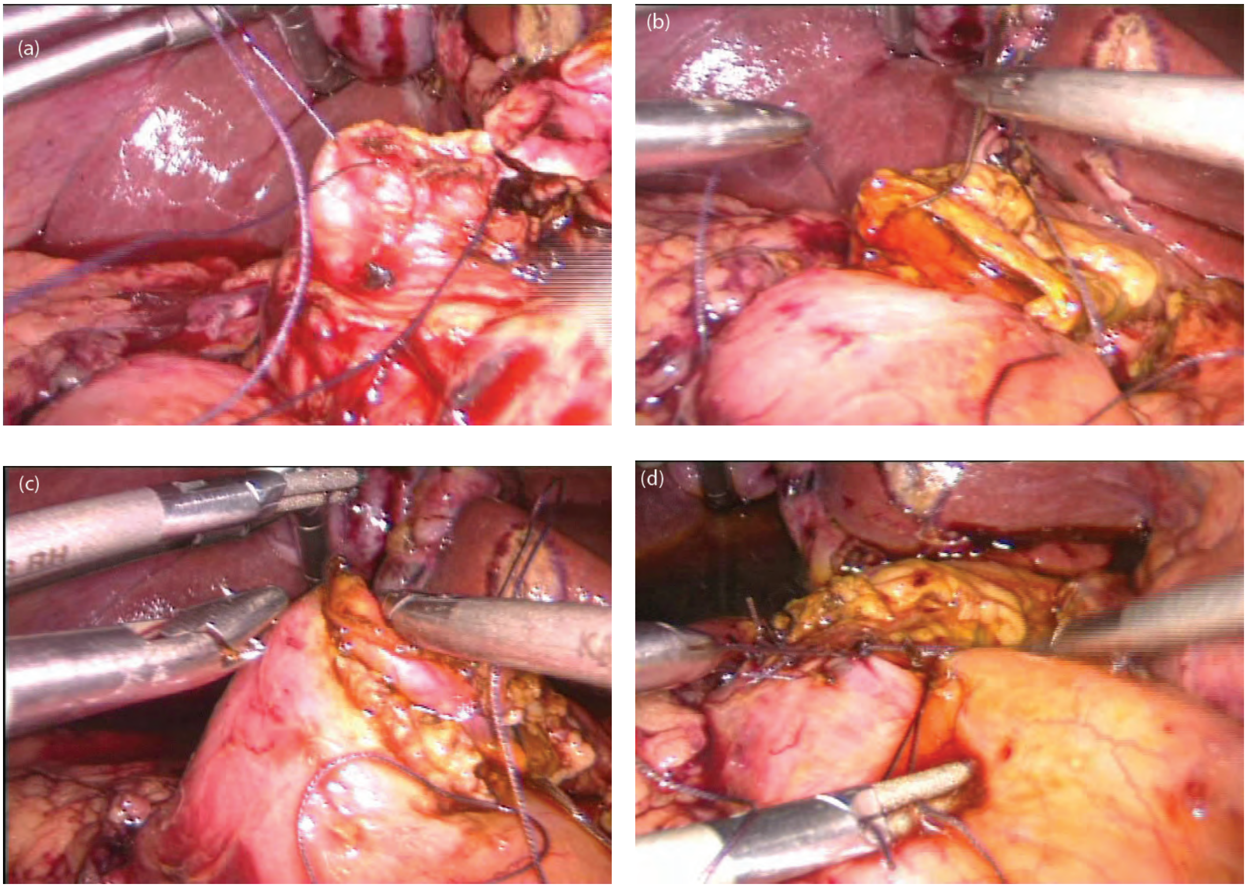


Figure 25.53 Laparoscopic transection choledochoduodenostomy for multiple ductal calculi and grossly dilated duct: (a) closure of the distal pancreatic end of the transected dilated common bile duct (CBD); (b) proximal end of the CBD approximated to the superior border of the duodenum; (c) completed running posterior anastomotic line; (d) interrupted anterior suture line using inverting deep seromuscular sutures.

the right and left hepatic ducts. Intrahepatic stones are categorized as:

- primary intrahepatic stones, which form within the intrahepatic ducts and are associated with intrahepatic duct strictures
- secondary, which form within the extrahepatic ducts but migrate subsequently to the intrahepatic ducts.

Primary intrahepatic duct stones are usually multiple and occur in association with pathological changes in the intrahepatic biliary tree: strictures, dilatations or angulation. The condition is also known as recurrent oriental cholangitis and has been recently reviewed.

Epidemiology

Primary intrahepatic stones are prevalent in East Asian countries (Taiwan, China, Japan and Korea) and rare in Western countries, the Middle East and Africa. In Taiwan, primary ductal calculi account for more than 50% of all cases of cholelithiasis and in Japan, 34%.

In countries of high prevalence, primary intrahepatic stones occur more frequently in rural areas, and this has been attributed to a low-protein, high-carbohydrate diet, although the reason for this dietary factor is not known. Some have suggested that a low-protein diet decreases glucuronolactone in bile, which is an inhibitor of β-glucuronidase. Glucuronolactone deficiency may thus lead to increased deconjugation of bilirubin diglucuronide with formation of calcium bilirubinate stones.

Pathology

The exact aetiology of primary intrahepatic calculi is not known, but several risk factors have been identified:

- infection by β-glucuronidase-producing bacteria
- parasitic infestations
- bile stasis and bile duct strictures
- congenital anomaly of bile ducts
- low-protein, high-carbohydrate diet
- low hygiene standards.

Bacterial infection of the biliary tree and bile stasis appear to be the most important as the prevalence of bacteria in the bile of patients with hepatolithiasis exceeds 90%. Furthermore, bile infection precedes formation of intrahepatic stones. Several bacterial species are responsible for the infection including *E. coli*, *Clostridium* spp. and *Bacteroides* spp. All exhibit β-glucuronidase enzyme activity, which is thought to be responsible for the hydrolysis of the soluble bilirubin glucuronide to the water-insoluble unconjugated bilirubin which then combines with ionized calcium and precipitates, forming calcium bilirubinate stones. Intrahepatic stones are usually associated with strictures and dilatations of the intrahepatic ducts.

Parasitic infestation, e.g. *A. lumbricoides* or *Clonorchis sinensis*, has also been implicated as a risk factor but this is unlikely for several reasons. In Taiwan, which has a very high incidence of intrahepatic lithiasis, infestation with *C. sinensis* is rare and in Japan, which has the second highest incidence, the prevalence of intrahepatic brown pigment stones has remained high despite the eradication of these parasites. Furthermore, the prevalence of gallstones in individuals of Chinese and Japanese descent born and living in the USA is no different from that of the general population.

Clinical features

Many patients with intrahepatic stones may remain asymptomatic for many years and often the condition is only discovered during routine investigation. Symptomatic patients can present with upper abdominal pain, occasional fever, rigors and, less frequently, jaundice. These symptoms are largely due to recurrent episodes of cholangitis. The laboratory findings include abnormal elevations of serum alkaline phosphatase, 5-nucleotidase and γ-GT. In addition, these patients have moderate elevations of serum transaminases and mild to moderate iron-deficiency anaemia. Untreated patients with longstanding disease with recurrent attacks of bacterial cholangitis can develop biliary cirrhosis with coagulation defects, low serum albumin and the development of ascites.

The important imaging modalities to establish the diagnosis include CT, MRCP and PTC. Ultrasound scanning is used for screening only as its diagnostic accuracy for documenting the full extent of hepatolithiasis is limited. In contrast, CT scanning provides detailed objective information on the number and location of the intrahepatic stones and the deformity/dilatation of intrahepatic ducts. MRCP is also useful in this respect. ERCP should be used with caution as, despite its established ability to accurately delineate the extrahepatic and intrahepatic ducts and detect intrahepatic stones, it can be complicated by severe cholangitis. PTC is very useful especially in the presence of strictures and has several advantages:

- It can be converted to percutaneous transhepatic biliary drainage (PTBD) in patients with obstruction and cholangitis.
- When established and mature, the PTBD tract can be dilated for the introduction of a flexible cholangioscope providing a percutaneous method of non-surgical treatment.

Treatment

Currently, two major treatment modalities are used with specific indications:

- surgical operative management
- percutaneous treatment.

Surgical management is still the mainstay of therapy. However, percutaneous transhepatic cholangioscopy removal gives good results and is emerging as the initial treatment in patients with bilateral disease.

Surgical treatment

This aims at both complete removal of stones and dilation of bile duct strictures to ensure adequate drainage with no bile stasis, essential for the prevention of recurrence and bacterial infections. When stones are located only in the left intrahepatic ducts and the affected left hepatic lobe is fibrotic and atrophic, left lateral hepatic segmentectomy or left hepatic lobectomy is the appropriate treatment. Likewise for disease located exclusively in the right intrahepatic ducts, right anterior or posterior segmentectomy may be necessary. Right hepatic lobectomy is however inadvisable because of surgical risks especially because the percutaneous cholangioscopic approach provides a much safer and effective treatment. In the past, an extended hepaticojejunostomy (Roux-en-Y) with a permanent cutaneous access was used for the removal of stones from the right hepatic ductal system. However, this has been largely replaced by the percutaneous transhepatic treatment.

When hepatolithiasis involves both intrahepatic ducts (more than 50% of patients), surgical treatment may be used for left lateral hepatic segmentectomy or left hepatic lobectomy and extended hepaticojejunostomy (Roux-en-Y) with permanent cutaneous access. The right intrahepatic stones are then removed percutaneously. Some would however disagree with this and recommend percutaneous treatment for bilateral stones as the first option before recourse to surgical treatment.

Percutaneous transhepatic approach

The percutaneous transhepatic approach has several advantages. Jaundice, when present, is relieved by PTBD. Provided the site of PTBD has been adequately selected, it will provide access for the insertion of the flexible endoscope used for the percutaneous treatment. This treatment is applicable to patients with multiple bilateral intrahepatic stones. Various modalities, e.g. electrohydraulic lithotripsy, laser lithotripsy, mechanical basket, can be introduced to remove or fragment stones. Strictures can be dilated by balloons or by bougienage and the passage of Yamakawa catheters through the tract.

However, percutaneous transhepatic cholangioscopic stone removal requires an experienced interventional endoscopist, preferably working with an interventional radiologist. The long-term results of percutaneous transhepatic cholangioscopy are similar to those of surgery. The most important factors affecting the recurrence are the presence of strictures and bile stasis.

Cholangitis

Acute bacterial cholangitis is a serious, life-threatening emergency caused by infection of an obstructed biliary tract. The systemic manifestations of the acute illness result from bacteraemia secondary to cholangiovenous reflux induced by the biliary hypertension ($<20\,\text{cmH}_2\text{O}$). The most common obstructing agent is an occluding stone in the common bile duct, followed by bile duct strictures (including sclerosing cholangitis) and tumours of the bile ducts, pancreatic head and periampullary lesions. Less commonly, cholangitis is secondary to bilioenteric anastomoses, spontaneous bilioenteric fistulas, cystic disease of the biliary tract and duodenal diverticula. Cholangitis may also occur following instrumentation of the biliary tract. Thus, it is encountered in 7% of patients undergoing ERCP. The risk factors for cholangitis following this investigation are the presence of fever before the procedure and malignant biliary obstruction. Thus, early decompression is indicated in these patients soon after the ERC. Cholangitis caused by contamination

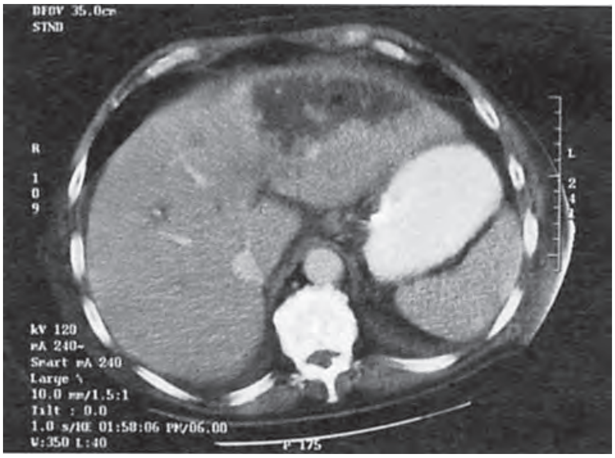


Figure 25.54 Intrahepatic segmental portal vein thrombosis with necrosis and abscess formation (central gas shadow) in a patient with severe cholangitis following preoperative stenting of a cholangiocarcinoma. The infected segmental necrosis developed after resection of the tumour. At operation gross purulent cholangitis was observed as the biliary tract was transected above the tumour. The condition proved fatal despite resection of segments II and III.

of the endoscopes (*Pseudomonas* spp.) has also been reported. Cholangitis occurs frequently (38%) as a complication of PTBD and is the main reason for abandoning preoperative decompression by this technique prior to surgical treatment in patients with severe obstructive jaundice. In Asia, recurrent pyogenic cholangitis is a frequent cause of recurrent bacterial cholangitis and is known as recurrent oriental cholangitis.

In severe cases of cholangitis there is neutrophil infiltration of the sinusoids and microabscess formation in the hepatic lobules. Portal thrombi and areas of hepatic necrosis are present in patients with severe disease accompanied by hypotension (Figure 25.54).

The infection is most commonly caused by Gram-negative organisms. The classical triad of symptoms consists of pain in the right hypochondrium, intermittent fever and jaundice (Charot's biliary fever), but the complete triad is present in only 70% of cases. Aside from toxicity, the high intermittent pyrexia is accompanied by severe rigors. The pain varies in intensity and may be severe. There is usually tenderness in the right hypochondrium, which, if marked, suggests the presence of abscess formation (honeycomb liver). The liver is often enlarged although this may be difficult to ascertain because of the abdominal tenderness. Nausea and vomiting are frequent accompaniments. Hypotension is present in patients with severe cholangitis, when renal failure is usually present.

Prompt and energetic treatment is mandatory. Resuscitative measures include crystalloid and colloid solutions. A blood culture is taken and systemic antibiotics are commenced (cephalosporin with metronidazole or piperacillin, or imipenem). The majority respond to this treatment but some (up to 30%) do not, and these patients require emergency biliary decompression. The patients who do not usually respond to conservative therapy include:

- females
- patients aged >50 years
- patients with acute renal failure
- patients with liver abscesses
- patients with bile duct stricture
- cirrhotic patients
- patients in whom the cholangitis is secondary to biliary instrumentation.

Biliary decompression may be accomplished surgically, endoscopically or by percutaneous techniques. For most patients with cholangitis and ductal calculi, endoscopic decompression by sphincterotomy and extraction of calculi is generally favoured, especially if the patient is elderly. If the calculi cannot be extracted a temporary pigtail stent draining the proximal biliary tree into the duodenum is inserted endoscopically. The alternative is surgical exploration with ductal clearance and insertion of a T-tube. Surgical intervention is the treatment of choice if the stones are large. Cholangitis secondary to biliary instrumentation often requires operative treatment. Percutaneous transhepatic drainage is useful in patients with cholangitis complicating strictures and malignant obstruction. The overall reported mortality of patients requiring urgent decompression for severe cholangitis is 15–20%.

Sclerosing cholangitis

This is an obscure disorder of uncertain aetiology, which results in a progressive fibrous obliteration of the biliary tract. The prevalence of primary sclerosing cholangitis (PSC) approximates to 20–40 per million population. Although sclerosing cholangitis has well-recognized histological, radiological and clinical features, there are no pathognomonic findings that reliably differentiate this disease from other hepatobiliary disorders. The distinction between primary and secondary types is no longer held to be valid. The term primary was formerly used to indicate no previous biliary surgery or biliary tract disease. Often, however, sclerosing cholangitis occurs as a secondary complication of inflammatory bowel disease, usually ulcerative colitis and, much less commonly, Crohn disease. The condition is currently regarded as an immune-complex disorder evoked by endotoxin–antibody complexes that have been identified in the peripheral blood of patients with inflammatory bowel disease.

The classification of the disorder is based on the extent of involvement of the biliary tree by the fibrous obliterative process (Table 25.3)

The disease results in extensive fibrosis which extends beyond the confines of the biliary ductal walls. Histologically, the fibrosis is concentric (onion shell) with patchy, chronic inflammatory infiltrate consisting of mononuclear cells and polymorphs (Figure 25.55). In addition, changes of cholestasis are seen. The gross fibrous thickening results in localized or multiple stricture formation. Although the ductal epithelium is frequently normal, it may become ulcerated and exhibit saccule formation. The

Table 25.3 Classification based on extent of involvement of the biliary tree

Type	Incidence (%)
Total diffuse	50
Localized hilar	25
Diffuse intrahepatic	10
Diffuse extrahepatic	10
Localized extrahepatic (distal)	5

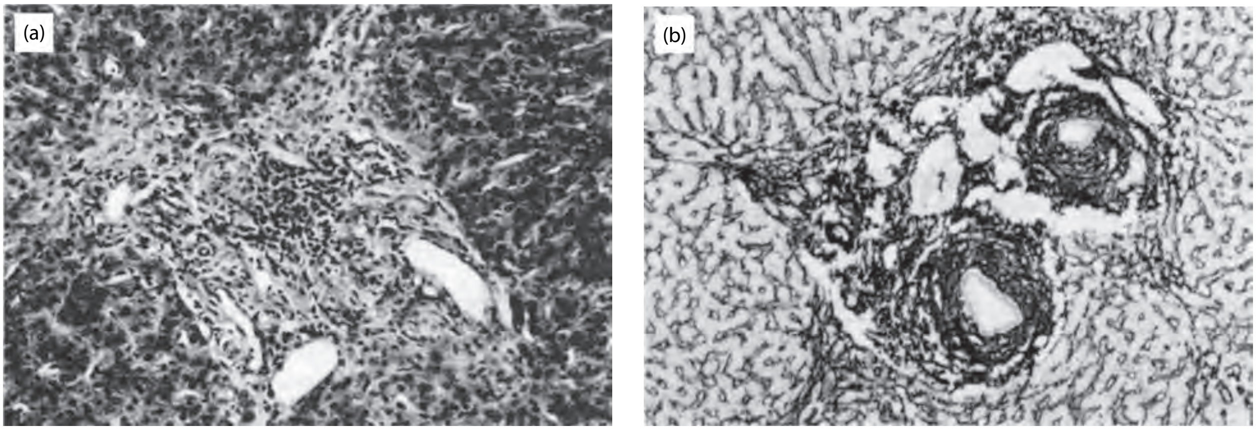


Figure 25.55 Sclerosing cholangitis: (a) histology of wedge liver biopsy (haematoxylin and eosin) showing narrowing of the intrahepatic ducts, fibrosis and cellular infiltrate; (b) reticulin stain of the same biopsy illustrating concentric deposition of fibrous tissue around the bile ducts.

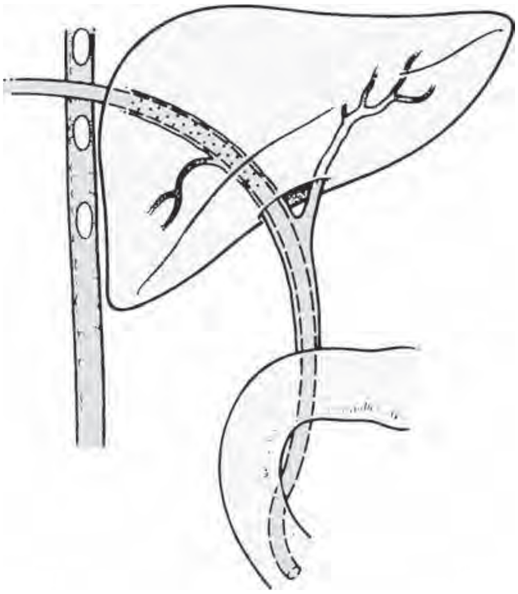


Figure 25.56 Technique of transhepatic stenting for total diffuse sclerosing cholangitis.

disease progresses invariably to cirrhosis and the development of portal hypertension.

Sclerosing cholangitis occurs more commonly in males (3:2) and usually presents in the fifth decade. The symptoms include vague ill health, asthenia, pain in the right hypochondrium, jaundice and itching, pyrexia and attacks of rigors. The liver is often palpable and tender. The liver function tests demonstrate a cholestatic picture and bilirubin is detected in the urine. The serum transaminases are mildly elevated. The majority of patients are hepatitis B surface antigen negative. Antimitochondrial, antismooth muscle and antinuclear antibodies are absent. Contrast radiological visualization shows pruning of the biliary tree (scanty ducts) and stricture formation which may be localized or diffuse (Figure 25.56). Globular dilatations (sacculations) are often seen in patients with diffuse disease. Differentiation from hilar disease and CC is often difficult on radiological grounds and may not be possible even after histological examination of biopsy specimens. In these patients, only the subsequent clinical outcome can identify the true diagnosis.

The majority of patients are symptomatic at the time of presentation. However, the correlation between symptoms and liver histology and disease progression is weak. In general, the median survival between presentation and the development of end-stage liver disease is 12 years. However, the rate of progression of the disease is very variable and is influenced by the frequency of episodes of bacterial cholangitis and the development of CC (central or peripheral), the diagnosis of which is often difficult in these patients. The most commonly used prognostic model is that of the Mayo Clinic:

- 0.535 $\log_e$ serum bilirubin (mg/dL)
- +0.486 histological stage
- +0.041 age (years)
- +0.705 if splenomegaly is present.

Unfortunately, there is no effective medical therapy for the condition. The pruritus may be controlled by cholestyramine. Episodes of cholangitis are managed by antibiotic therapy. Surgical intervention may be considered when adequate control of symptoms is not achieved by medical therapy, the specific indications being progressive jaundice and recurrent cholangitis. Surgical treatment gives best results for localized hilar disease. The hilar bifurcation is accessed through an anterior segmentectomy IV and a Roux-en-Y hepaticojejunostomy to the right and left hepatic ducts performed proximal to the stricture. Direct surgical intervention on the biliary tract for predominantly hilar disease is contraindicated in the presence of cirrhosis as the results are poor and the mortality is high.

For diffuse disease, intraoperative dilatation of the intrahepatic biliary tree via a choledochotomy using both metal and balloon dilators is followed by a large silicone stent introduced transhepatically down the bile duct into the duodenum. The stent is exteriorized through the right lowest intercostal space in the anterior axillary line. Daily irrigation with heparinized saline and prolonged antibiotic therapy are essential components of the postoperative management. The stent is left in place for at least 12 months and progress is assessed by repeat cholangiograms carried out through the stent (Figure 25.57). Replacement of the stent may be necessary if it becomes blocked by encrustation with calcium bilirubinate. Both percutaneous transhepatic and endoscopic

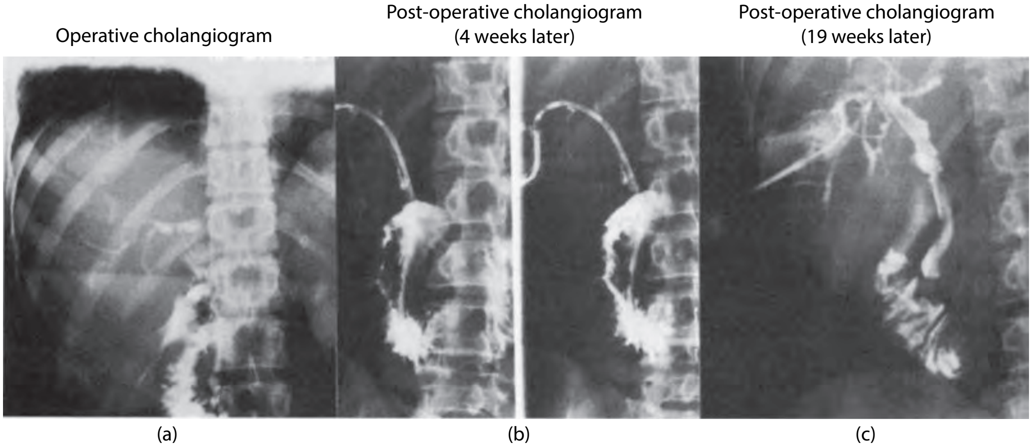

Operative cholangiogram
Post-operative cholangiogram (4 weeks later)
Post-operative cholangiogram (19 weeks later)
(a)
(b)
(c)

Figure 25.57 Marked radiological improvement in a patient with sclerosing cholangitis following prolonged stenting of the intra- and extrahepatic biliary tract. The patient showed a dramatic clinical and biochemical improvement. (a) Operative cholangiogram; (b) postoperative cholangiogram (4 weeks later); (c) postoperative cholangiogram (19 weeks later).

balloon dilatation and stent insertion have been employed to deal with strictures in patients with sclerosing cholangitis. Generally the results have been inferior to those achieved in patients with iatrogenic strictures following bile duct damage.

However, the results of surgical treatment for diffuse disease have been disappointing and these patients are better served by hepatic transplantation. The optimal timing of hepatic transplantation for PSC has not been defined and thus early referral to a transplantation centre is currently recommended. Certainly a Mayo score of ≥5 or Child C constitute definite indications for transplantation. The problem in these cases is the exclusion of a superadded CC, the presence of which is accompanied by a dismal prognosis and is a contraindication to transplantation.

Biliary disorders in AIDS

Biliary disorders seen in AIDS patients are AIDS cholangiopathy (AC) and acalculous cholecystitis (ACC). Since the advent of highly active antiretroviral therapy (HAART) the incidence of AIDS-associated biliary disease has declined. The two diseases can occur concurrently.

AIDS cholangiopathy

The average age of AC diagnosis is 37 years with a range of 21–58 years. AC is regarded as an infectious form of sclerosing cholangitis. *Cryptosporidium parvum* is the most common organism associated with AC but other pathogens include cytomegalovirus, *Giardia lamblia*, *Mycobacterium avium-intracellulare*, *Cyclospora cayetanensis*, *Isospora belli*, *Histoplasma capsulatum* and *microsporidia*.

Clinical features

The clinical features are the same regardless of the opportunistic organism involved. AC presents with right upper quadrant and/or mid-epigastric abdominal pain in 90% of patients. Other symptoms include fever, nausea, vomiting and diarrhoea, which is due to the small bowel involvement. In addition, 75% of patients experience significant weight loss, and jaundice associated with pruritus is present in 10%. Patients who develop AC typically have a CD4 count <100/mm^3 at presentation. LFTs show a cholestatic pattern (elevated γ-glutamyl transferase and an increased alkaline phosphatase of 700–800 IU/L). Most patients also have a mild increase in transaminases and total bilirubin. However, nearly 20% have normal liver function tests. The most common findings on an abdominal ultrasound include common bile duct dilatation in 65–75% of cases, and thickening of the common bile duct wall in 20–40%. In addition, an echogenic nodule is often present at the distal end of the common bile duct caused by oedema of the papilla of Vater. ERCP is the definitive test and is used therapeutically for symptomatic relief in patients with papillary stenosis. ERCP identifies four common patterns of cholangiographic abnormalities:

- papillary stenosis (36%)
- intra/extrahepatic sclerosing cholangitis (≤15%)
- papillary stenosis combined with sclerosing cholangitis in 44%
- isolated long extrahepatic duct stenosis or stricture in 5%.

Treatment

Papillary stenosis is treated by endoscopic sphincterotomy, which provides symptomatic relief of pain in 87–100% of patients, although it does not improve survival. Despite sphincterotomy, the serum alkaline phosphatase often continues to rise, indicative of progression of intrahepatic disease. Biliary stents and balloon dilation of strictures are also used in some patients. Ursodeoxycholic acid has also been used as an additional treatment and can result in symptomatic improvement accompanied by a fall in serum alkaline phosphatase. Combination antiretroviral therapy is used to suppress viral replication and usually leads to elevation in CD4 count. The reported median survival of 30–35 months is attributable to HAART. A normal or slightly elevated serum alkaline phosphatase is associated with a better outcome, whereas high levels (>1000 IU/L) are associated with increased mortality.

Acalculous acute cholecystitis

The actual incidence of ACC in HIV patients is unknown. Patients with ACC usually have end-stage HIV disease with low CD4 counts, coexisting opportunistic infections and significant malnutrition. The organisms implicated are cytomegalovirus, *Cryptosporidium*, microsporidia, *Salmonella enteritis*, *Pneumocystis carinii*, *Campylobacter*, *I. belli* and *Candida albicans*. However, in a substantial cohort (50%) of patients, no definite pathogen is identified on microbiological studies. ACC can also be caused by obstruction of the cystic duct, due to Kaposi's sarcoma.

Clinical features

The diagnosis of ACC is clinical and should be considered when HIV patients complain of right upper quadrant pain with right upper quadrant (90%) tenderness, unexplained fever or signs of peritonitis. Other features include nausea (80%), diarrhoea (60%) and weight loss (40%). The serum bilirubin is usually normal but the serum alkaline phosphatase is elevated. The serum transaminases are either normal, or mildly elevated. Ultrasound is used to establish the diagnosis in the first place (pericholecystic fluid, intraluminal gas, gallbladder wall thickening ≥3.5 mm, sloughed mucosal membrane, echogenic bile, subserosal oedema and distended gallbladder in the absence of gallstones). The diagnostic criteria of CT are similar but it may also demonstrate other intra-abdominal pathologies.

Treatment

The treatment is cholecystectomy. Percutaneous CT or ultrasound-guided cholecystostomy is performed for gallbladder decompression in ACC patients who are extremely ill or pose a high surgical risk. It can serve as definitive therapy or a temporizing procedure until the patient's condition improves to enable safe cholecystectomy. If cholecystectomy is performed the resected specimen is sent for both microbiological and histological examination. Postoperative mortality after cholecystectomy in HIV-seropositive patients ranges from 2% to 22%. The deaths are the result of complications. The CD4 lymphocyte count does not influence postoperative mortality.

Duodenal diverticula

The prevalence of periampullary duodenal diverticula in patients undergoing ERCP is 12.5% and appears to increase with age. Their importance is twofold. In the first instance, they predispose to bacterial infection. Thus patients with duodenal diverticula have a higher incidence of infected bile than normal subjects and the bacterial count of their duodenal contents is elevated (with proliferation of *Enterobacteriaceae* spp.) compared with patients without diverticula. Second, duodenal diverticula increase the difficulty of cannulation of the papilla during ERCP and thus the failure rate of this procedure, especially when the opening of the duodenal papilla lies inside a diverticulum.

The duodenum is the second most common site of gastrointestinal diverticula. Duodenal diverticula occur in approximately 20% of adult individuals but the vast majority are asymptomatic and discovered by chance during investigation of the upper gastrointestinal tract. Duodenal diverticula rarely cause symptoms, which are usually non-specific unless they develop complications. Duodenal diverticula are nowadays classified as primary (the vast majority) or secondary to another disease such as chronic duodenal ulceration, i.e. prestenotic diverticulum. In clinical practice, duodenal diverticula are discovered mainly in middle age (50–60 years). Their reported incidence varies with the investigating modality used, being low in upper gastrointestinal contrast studies (up to 6%) and much higher in reported ERCP series (up to 23%), which approximates to the incidence reported in autopsy series (averages 22%). More than 95% of duodenal diverticula project from the inner (pancreatic) border of the duodenal curve of the second to fourth parts of the duodenum, with the second part being the commonest site (85%). Duodenal diverticula may be single or multiple and are spherical unless very large, when they assume a flask shape. There are various classifications which are rather confusing (ampullary, periampullary, etc.). The current classifications based on CT or MRI collectively address these as juxtapapillary duodenal diverticula (JPDD), and undoubtedly these are the surgically important in view of their pathogenetic potential. The CT-based classification of JPDD proposed by Weisner *et al.* is shown in Box 25.2.

A similar but more anatomical detailed classification based on MRI contrast study has been proposed by Morita *et al.* and is illustrated in Figure 25.58.

BOX 25.2 Radiological (CT) classification of juxtapapillary duodenal diveticula

- Type 1: Diverticulum is located ventral to the ampulla of Vater extending into the pancreas at the site of fusion of the dorsal and ventral anlage of the pancreas (commonest)
- Type II: Diverticulum is dorsal to the ampulla of Vater (second commonest)
- Type III: Bilobulated diverticulum extending above and below the ampulla of Vater (rare)
- Type IV: Diverticulum ventral to the minor papilla (least common)

Aetiology and pathology

Except for a rare small subgroup in childhood, duodenal diverticula are acquired and consist of a sac of mucosal/mucosal + submucosal layers herniated through muscular defects in the duodenal wall although the exact mechanism for this herniation is not known. One hypothesis is that herniation occurs through localized mural weakness caused by pancreatic ducts/blood vessels traversing the duodenal wall. Once the diverticulum forms, it slowly increases in size with time. A large duodenal diverticulum may lie alongside or extend behind the pancreas and occasionally common bile duct, and the duct of Wirsung may open into the duodenal diverticula.

Pathologically all large duodenal diverticula lack a muscular coat. The distinction between true (congenital) and false diverticula depending on the presence or absence of muscle coat is no longer tenable. In the first instance, there is no evidence to prove that certain pouches are congenital and that others are acquired. Second, in duodenal diverticula the amount of muscle fibres in the pouch depends on its size. In the majority of duodenal diverticula, the circular and longitudinal muscle coats of duodenum are missing, the wall of the diverticulum being composed of mucosa and muscularis mucosae.

Clinical features

The vast majority of duodenal diverticula remain asymptomatic. Clinical symptoms when present are non-specific and consist of abdominal discomfort usually located in the epigastrium, right upper quadrant or umbilical region, which is made worse

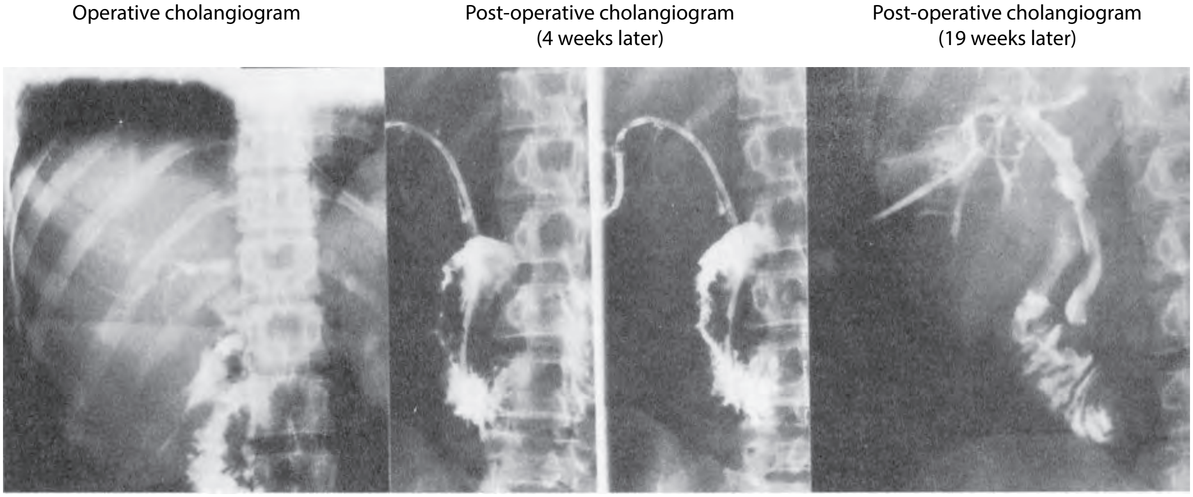

Operative cholangiogram
Post-operative cholangiogram (4 weeks later)
Post-operative cholangiogram (19 weeks later)

Figure 25.58 Schematic classification of juxtapapillary duodenal diverticula (JPDD) with respect to its location to the papilla: type 1, no JPDD; type 2, JPDD with the papilla opening inside it; type 3, JPDD with the papilla located on its dorsal side; type 4, JPDD with the papilla located on its ventral side; type 5, JPDD with the papilla located between two diverticular openings (bilocular). (From Morita M, Ueno E, Masukawa A, *et al. Jpn J Radiol* 2009;**27**:423–9, with permission.)

or precipitated by eating and relieved by vomiting or belching. These symptoms may be caused by delayed emptying of duodenal diverticula or pressure on the common bile or pancreatic duct, or subacute obstruction of the duodenum. Others attribute these symptoms to inflammatory changes following stasis in the duodenal diverticula. Some duodenal diverticula produce peptic ulcer-like dyspepsia. Intermittent diarrhoea and constipation, weight loss and steatorrhoea may also occur in some patients. Some patients with duodenal diverticula located near the ampulla of Vater may simulate biliary tract disease and when inflamed (peri-Vaterian diverticulitis) may give rise to obstructive jaundice. There is an established association between duodenal diverticula and brown pigment stones which is not fully understood although colonization with β-glucuronidase-producing bacteria of the common bile duct has been proposed. Since choledocholithiasis is an established pathogenetic factor in the aetiology of acute biliary pancreatitis, the increased incidence of pancreatitis (acute and chronic) in patients with duodenal diverticula has been attributed to the high prevalence of ductal calculi.

JPDD are not uncommonly associated with biliary or pancreatic disease and can cause cholangitis without choledocholithiasis, known as *Lemmel syndrome*. JPDD can cause difficult cannulation of the common bile duct and complications during ERCP. JPDD can also be mistaken for cystic neoplasms of the pancreas. The complications of JPDD are caused by:

- pressure on adjacent structures (mechanical compression): jaundice, cholangitis, pancreatitis
- diverticulitis
- ulceration
- acute obstruction
- perforation
- duodenocolic, gastrojejunocolic fistula (rare)
- distal small bowel intraluminal obstruction.

Diverticulitis ensues as a result of stasis and bacterial proliferation in duodenal diverticula, which cannot drain their contents and may lead to perforation with either localized abscess formation or generalized peritonitis. Perforation may also be caused by enteroliths or foreign ingested material. The perforation may be acute (majority of cases) or chronic leading to a bile-stained phlegmon in the paraduodenal region, retroperitoneal oedema or a right subhepatic abscess. Ulceration can occur because of the presence of ectopic gastric mucosa or inflammation. Diverticular ulceration may give rise to occult bleeding or very rarely massive gastrointestinal haemorrhage leading to hypovolaemic shock due to erosion of a large mesenteric vessel. Duodenocolic and gastrojejunocolic fistula formation secondary to inflammation in duodenal diverticula have been reported but are rare. Pancreatic and/or biliary fistula do not arise *de novo* but may develop as complications after surgery. Intestinal obstruction as a result of blockage by large enteroliths derived from duodenal diverticula has been reported.

A number of disorders are associated although there is no firm evidence for a causal relationship except for duodenal ulcer. The associated disorders include:

- duodenal ulcer
- hiatus hernia
- jejunal and ileal diverticula in 4–13% of all patients with duodenal diverticula
- higher than expected incidence of coexisting congenital anomalies: annular pancreas, choledochocele, omphalocele, hypoplastic kidney, situs inversus, malrotation, Ladd band, portal vein anomalies and Down syndrome.

The diagnosis may be established by enteroclysis, upper gastrointestinal endoscopy and ERCP. CT or MRI are performed when perforation is suspected and may demonstrate a soft-tissue mass between the duodenum and the pancreatic head containing an air–fluid level, duodenal wall thickening and surrounding free air. MRI facilitates precise delineation of the complicated duodenal diverticula while MRCP allows assessment of the effects of duodenal diverticula on the biliary and pancreatic duct.

Treatment

Asymptomatic duodenal diverticula do not require treatment and only a minority of diagnosed duodenal diverticula (up to 5%) require surgery because of symptoms or complications. Acute obstruction of a diverticulum is treated by endoscopic clearance. Operative treatment requires complete Kocherization of the duodenum. The diverticulum is then dissected free and opened to establish its relation to the ampulla of Vater before its excision. The supraduodenal common bile duct is opened and a T-tube inserted. If the ampulla of papilla is inside the duodenal diverticula, some advocate a choledochoduodenostomy or choledochojejunostomy rather than direct surgery on the duodenal diverticula. However, this approach may not deal with problems arising from involvement of the pancreatic duct. In cases where excision is deemed hazardous, exclusion by division of duodenum some 2.0 cm distal to the pylorus with drainage by a Roux-en-Y duodenojejunostomy may be performed.

The surgical treatment for perforated duodenal diverticula carries a high postoperative mortality and morbidity from iatrogenic injury to the ampulla of Vater leading to impaired drainage of bile or pancreatic drainage with the development of jaundice, pancreatitis or duodenal fistula.

Haemobilia

Major (profuse) haemobilia is rare; when it occurs spontaneously, it is caused by rupture of an intrahepatic aneurysm. Nowadays, significant haemobilia is most commonly encountered as a complication of percutaneous radiological interventions on the liver and following hepatic trauma. The vascular injury within the hepatic parenchyma caused by transhepatic radiological interventions results in an arteriovenous fistula or pseudoaneurysm or direct vascular–biliary connection. If the bleeding is marked, blood clots with biliary colic complicate the clinical picture. Minor haemobilia may also be caused by stones, primary hepatic tumours including angiomas, gallbladder and bile duct polyps, parasitic infestations, severe cholangitis and coagulation disorders. Haemobilia may complicate ESWL for ductal calculi.

In descending order of frequency the causes of haemobilia are:

- iatrogenic trauma
- accidental trauma
- gallstones
- acalculous inflammation
- vascular conditions
- neoplastic disease
- others.

Post-traumatic haemobilia is the commonest cause (55–80%) and may complicate both blunt and penetrating abdominal injuries and percutaneous interventions used commonly nowadays in the management of hepatobiliary disorders. Iatrogenic haemobilia may also follow operations on the liver and biliary tract. Upper gastrointestinal bleeding due to gallstones accounts for 9–15% of all reported cases of haemobilia but is invariably minor. The bleeding results from mucosal erosions of the gallbladder mucosa or haemorrhagic necrosis in severe cholecystitis. Haemobilia is also well documented in acalculous inflammatory conditions of the biliary tract, with the incidence being higher in regions where parasitic infestations are endemic. The leading cause in this category is ascariasis. Haemobilia may also complicate other conditions: hepatic abscess and cholangitis.

A primary vascular cause is present in 11–15% of reported cases of haemobilia, e.g. aneurysms of the hepatic artery and portal hypertension although the latter is nowadays regarded more as a predisposing rather than a primary cause. The bleeding from vascular causes tends to be massive and life-threatening. Neoplastic disease is a relatively rare cause and accounts for 6% of reported cases of haemobilia, the most common tumour being hepatocellular carcinoma followed by haemangioma, CC and carcinoma of the gallbladder.

Clinical features

In postoperative patients with biliary drains, the diagnosis may be obvious with the appearance of blood and clots in the drain. Otherwise the diagnosis may not be readily apparent. The clinical features include biliary colic (70%), jaundice (60%) and gastrointestinal bleeding, which may be occult or massive. The diagnosis may be established by endoscopy, angiography and CT (for hepatic injuries) or during surgery by intraoperative cholangiography.

Treatment

Expectant therapy with observation and close monitoring is indicated when haemobilia occurs as a complication of liver biopsy and percutaneous transhepatic cholangiography as these usually subside with supportive measures. Endoscopic techniques include nasobiliary drainage, endoscopic sphincterotomy and laser/electrocoagulation. The therapeutic modalities used for persistent haemobilia from any source are:

- angiography with embolization
- surgical intervention
- electrocoagulation
- photocoagulation.

Angiography is the most effective therapy for controlling bleeding from intrahepatic causes and has a reported efficacy of 95%. Surgical treatment is indicated for extrahepatic causes: bleeding from the gallbladder mucosa (stones, acalculous cholecystitis, etc.), bile duct and gallbladder cancers. Surgical treatment is also indicated for haemobilia after blunt trauma for which limited debridement, vessel ligation and drainage are important

Bilioenteric fistulas

The various types and causation of biliary fistulas are shown in Table 25.4.

Spontaneous external biliary fistulas are exceedingly rare and the few reported cases have been instances of neglected empyema of the gallbladder or extensive carcinoma of the gallbladder invading the abdominal wall. The vast majority of

Table 25.4 Biliary fistulas

Category	Type	Causation
External		Trauma and operative injuries therapeutic (T-tube, stents, cholecystostomy)
Internal	Bilioenteric	
	Cholecystoduodenal	Gallstones
	Cholecystocolic	Gallstones, Ca
	Cholecystogastric	Gallstones, Ca, peptic ulceration
	Choledochoduodenal	Ductal calculi, iatrogenic, duodenal ulcer, Ca
	Biliobilial	
	Cholecystocholedochal	Gallstones (Mirizzi syndrome)
	Others	Trauma, operative injuries, liver abscesses/hydatid, subphrenic abscesses
	Broncho/pleurobilial	
	Cholecystorenal	Gallstones

external biliary fistulas occur in the postoperative period and may result from the following:

- leakage of bile from a slipped cystic duct ligature or cut accessory bile duct
- trauma to the extrahepatic biliary tree during cholecystectomy, gastric surgery or pancreatectomy
- dislodged T-tube after common bile duct exploration
- leakage from bilioenteric anastomosis
- hepatic resections.

Leakage of bile after removal of a T-tube is short-lived and requires investigation by ERCP if it persists beyond 2–3 days. Other external biliary fistulas may follow blunt or penetrating hepatic trauma. The external biliary fistula usually occurs after surgical treatment of the hepatic injury and is then often accompanied by sepsis. External biliary fistulas do not result in skin excoriation but may cause significant fluid and electrolyte depletion if the output is high and prolonged. They are not usually accompanied by systemic manifestations unless sepsis is present. Abdominal tenderness and rebound indicates the concomitant presence of bile in the peritoneal cavity. Postoperative external biliary fistulas occurring in association with jaundice indicate bile duct trauma or a missed obstructive lesion of the biliary tract. These patients require urgent investigation with ultrasonography/CT and ERCP. Otherwise, a conservative management regimen is adopted as the biliary fistula usually closes spontaneously. Persistence of the fistula beyond a reasonable period (7–10 days), the development of jaundice, pyrexia or deterioration of the patient's condition are indications for urgent reassessment and investigation.

Internal fistulas are usually spontaneous and arise from chronic or acute perforation of the gallbladder into an adjacent organ. Others are due to malignant infiltration arising from or involving the gallbladder, e.g. carcinoma of the hepatic flexure, duodenum or gallbladder. The symptoms of the non-malignant internal fistulas involving the gallbladder are similar to those of chronic cholecystitis but jaundice and cholangitis are more common and radiology of the abdomen shows gas or barium in the biliary tree. The most frequent of the internal fistulas is the cholecystoduodenal fistula followed by cholecystocolic and cholecystogastric fistulas. The Mirizzi syndrome refers to a condition characterized by obstructive jaundice caused by a stone impacted in the neck of the gallbladder which compresses the common hepatic duct and which eventually ulcerates through into the common hepatic duct causing a cholecystocholedochal fistula (Figure 25.59).

A fistulous tract between the lower end of the bile duct and the duodenum (choledochoduodenal fistula) may arise spontaneously (secondary to ductal calculi or chronic duodenal ulcer) or be the result of iatrogenic injury from ill-advised

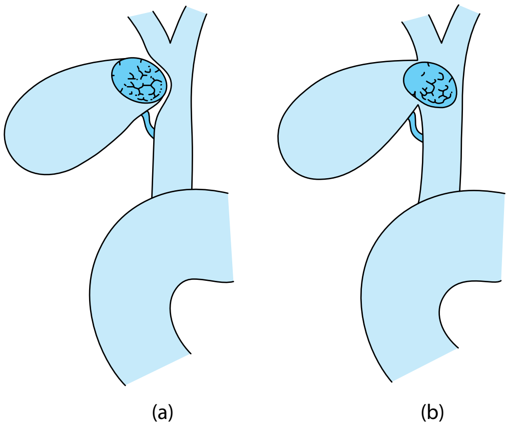


Figure 25.59 The Mirizzi syndrome: (a) a stone impacted in the neck/Hartmann's pouch causes extrinsic compression of the common hepatic duct followed by (b) fistula formation between the gallbladder and the common hepatic duct.

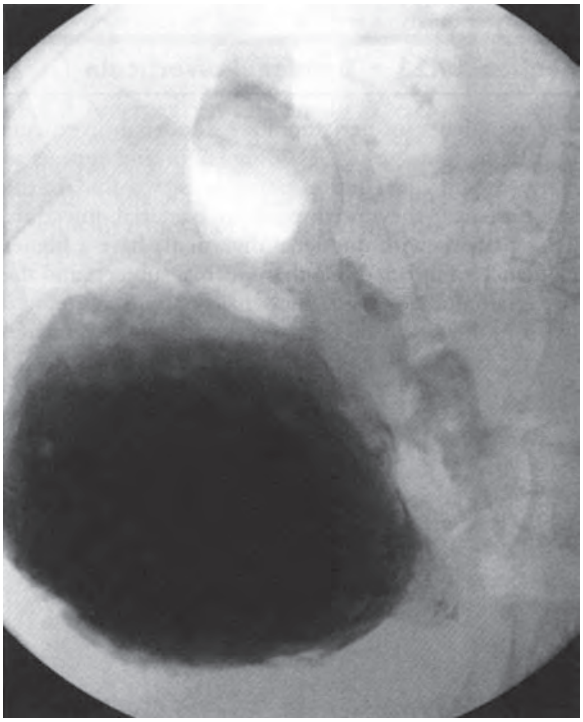


Figure 25.60 Contrast study of a large hepatic abscess with communication with the bronchial tree. At operation the abscess had burrowed through the diaphragm to involve the lower lobe of the lung which required resection.

probing of the Vaterian segment of the common bile duct during biliary surgery. Biliopleural and bronchobilial fistulas are usually the result of hepatic abscesses (Figure 25.60) and hydatid disease of the liver, although some follow hepatic injuries complicated by the development of subphrenic abscesses.

The treatment of bilioenteric fistulas due to gallstone disease consists of a cholecystectomy and closure of the fistulous communication. Exploration of the common bile duct is frequently necessary and is dictated by the findings at peroperative cholangiography. Unless the Mirizzi syndrome is recognized at operation, damage to the common hepatic duct is inevitable. The surgical treatment of this condition entails leaving a small cuff of gallbladder wall, which is used to close the fistulous opening. The common bile duct is explored (if necessary) through a choledochotomy lower down. The management of bronchobiliary fistulas consists of adequate drainage of the underlying hepatic/subphrenic abscess and decompression of the biliary tract when necessary.

Gallstone ileus

This condition, which characteristically affects elderly people, is due to intraluminal intestinal obstruction by a large gallstone that enters the intestinal tract subsequent to the establishment of a fistula, usually between the gallbladder and the duodenum and, less commonly, the gallbladder and the colon. Rarely gallstone ileus may occur as a complication of endoscopic sphincterotomy with stone extraction. Naturally occurring gallstone ileus occurs in 2% of patients with gallstone disease and, in some reports, accounts for up to 20% of mechanical intestinal obstruction in elderly people.

The patient, who may give a history of gallbladder disease, presents with acute intestinal obstruction, which in the vast majority of cases affects the small bowel, colonic obstruction being distinctly uncommon. Characteristically, the level of the obstruction is changing until the stone becomes firmly impacted, usually in the terminal ileum (70%), as this is the narrowest part of the intestinal tract, and much less commonly in the duodenum. Colonic obstruction due to impaction in the colon is the result of a cholecystocolic fistula.

The condition should be suspected in the elderly patient with mechanical intestinal obstruction in the absence of the more common causes of this condition. It can be diagnosed preoperatively if gas can be demonstrated in the biliary tract or the gallstone is visualized, usually in the right iliac fossa (Figure 25.61).

The treatment requires emergency surgical intervention in all patients. The operative management depends on the findings and the general condition of the patient. In the elderly and frail patient with ileal obstruction, removal of the impacted calculus through a small enterotomy is performed and the cholecystoduodenal fistula is dealt with at a subsequent operation. A one-stage enterolithotomy with cholecystectomy and closure of the duodenal fistula can be performed in patients who, despite their age, are considered fit enough for this procedure. The treatment of patients with colonic obstruction and a cholecystocolic fistula consists of removal of the calculus through a colotomy, cholecystostomy (cholecystectomy if the patient is fit) and exteriorization of the colotomy as a temporary proximal (diverting) colostomy.

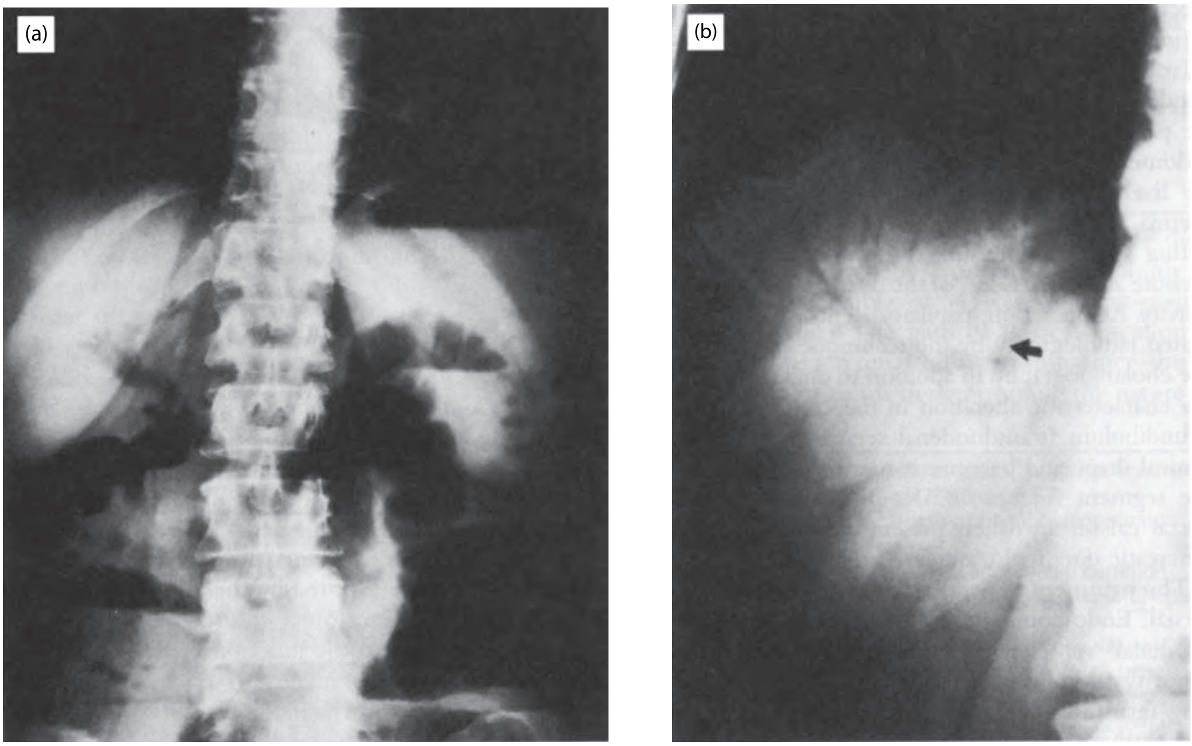


Figure 25.61 (a) Plain radiograph of the abdomen – gas outlines the common bile duct. The patient presented with cholangitis and gallstone ileus. A spontaneous cholecystoduodenal fistula was found at operation. (b) Gas outlining the right hepatic duct. The patient had an empyema of the gallbladder and a cholecystocolic fistula due to a carcinoma of the hepatic flexure.

Postcholecystectomy syndromes

These refer to the persistence of symptoms referable to the biliary tract after cholecystectomy. As currently defined, the syndromes exclude those patients whose symptoms are due to organic disease outside the biliary tract. These constitute a significant percentage of patients with persistent symptoms after cholecystectomy and they are usually a reflection of failure of proper evaluation and investigation of patients prior to the cholecystectomy.

The reported incidence of postcholecystectomy syndromes varies widely and correlates with the duration of follow-up. There is a female preponderance, particularly in the 40–50 years age group. A careful evaluation and a full investigation of the biliary tract including an MRCP is advisable in all patients with persistence or recurrence of symptoms after cholecystectomy. The common causes of postcholecystectomy syndromes are:

- retained or recurrent calculi
- gallbladder/cystic duct remnants
- bile duct strictures and other unrecognized iatrogenic injuries (choledochoduodenal fistula)
- papillary stenosis and biliary dyskinesia.

Persistent or recurrent symptoms after cholecystostomy are common and are one of the reasons for subsequent cholecystectomy in all patients who are considered fit for surgery. Controversy still exists regarding the role of a 'long cystic duct remnant' as a cause of persistent symptoms after cholecystectomy. There are undoubtedly patients in whom a dilated long cystic duct remnant containing stones is demonstrated on investigation, and its removal together with the stones results in sustained symptomatic improvement. However, these cases are few and far between and at present there is no evidence to incriminate an otherwise normal long cystic duct remnant as one of the important causes of postcholecystectomy syndrome.

Papillary stenosis (also known as choledochoduodenal junctional stenosis) is nowadays regarded as a rare but definite entity, which results from fibrosis or fibromuscular hyperplasia of the sphincter of Oddi. An associated duodenal diverticulum is common, and cannulation of the papilla is difficult. In addition to pain in the upper abdomen, the patient may exhibit slight abnormalities in LFTs, including mild hyperbilirubinaemia and elevated alkaline phosphatase activity. The resting sphincter pressure is elevated as is the passage pressure and there is loss of the normal phasic sphincteric activity. At operation, papillary stenosis is best demonstrated radiologically by the technique of contact selective cholangiography. In addition to duct dilatation, there is a characteristic alteration in the configuration of the infundibulum (transduodenal segment), which loses its conical shape and becomes wider than the intrapancreatic segment (Figure 25.62). Biliary sludge and small ductal calculi are often present. Reflux into a dilated pancreatic duct is also observed in some cases.

The treatment of papillary stenosis is equally controversial. Endoscopic sphincterotomy or surgical transduodenal sphincteroplasty is recommended by the majority although there has not been an adequate long-term assessment of these procedures for this rare and elusive condition. In addition to the sphincteroplasty, Moody advocates the excision of the septum between the pancreatic duct and the bile duct in patients with chronic pain which he maintains is of pancreatic origin. An alternative surgical treatment for papillary stenosis favoured by the author is transection choledochoduodenostomy with reimplantation of the mobilized duct into the junction of the first with the second part of the duodenum (Figure 25.63).

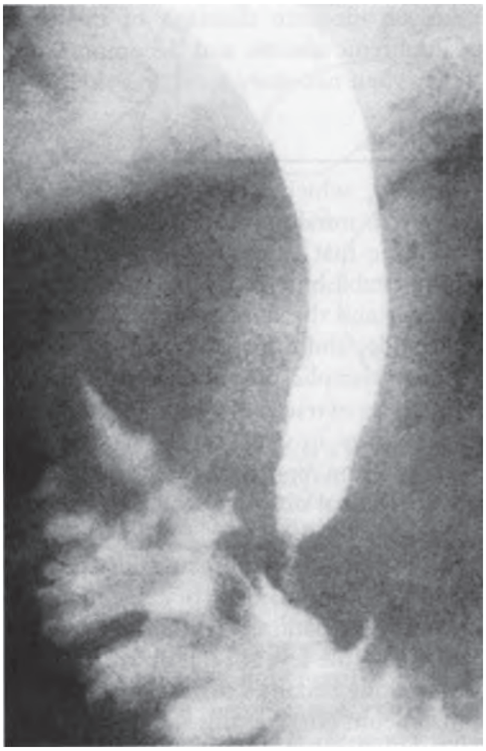


Figure 25.62 Operative contact selective cholangiogram demonstrating papillary stenosis. There is minimal dilatation of the bile duct and the infundibulum becomes globular. Biliary sludge/small ductal calculi are often present.

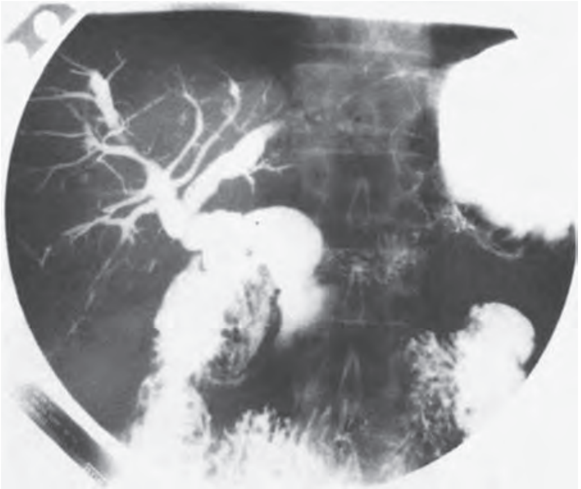


Figure 25.63 Barium meal contrast study after transection choledochoduodenostomy performed for papillary stenosis 3 years previously.

The term biliary dyskinesia is used to denote those patients who have persistent pain after cholecystectomy and no other abnormality on physical examination and routine testing but who exhibit the following abnormalities during ERCP manometry:

- elevated resting pressure
- tachyarrhythmia (increased phasic activity of the sphincter)
- retrograde contractions of the sphincter
- paradoxical response to CCK.

Treatment with endoscopic sphincterotomy has been advocated for these patients but the efficacy of this in the long-term relief of symptoms remains to be ascertained.

Benign bile duct strictures

The causes of benign bile duct strictures are:

- operative bile duct injury (iatrogenic)
- penetrating and non-penetrating abdominal injuries
- chronic duodenal ulcer
- chronic pancreatitis
- recurrent pyogenic cholangitis and parasitic infestations
- sclerosing cholangitis.

In the clinical context, benign strictures of the extrahepatic bile ducts do not exhibit a benign course since they are always attended by significant symptoms and serious complications that are life-threatening in both the short and the long term and carry a definite mortality. Thus they pose serious health and economic problems and may expose the surgeon to expensive medicolegal litigation. In addition, they increase substantially the economical burden to the patient, hospital and the community, and some have needed hepatic transplantation for survival. The costs of repair of cholecystectomy-related bile duct injuries is high, varying from 4.5 to 26 times the cost of the uncomplicated procedure. In one reported series, patients with LC-related bile duct injuries were billed a mean of US$51 411 for the remedial surgery and incurred an average of 32 days of inpatient hospital stay. Both the costs of treatment and the outcome (mortality) of bile duct injury are related to early recognition. Thus in one large reported series, bile duct injuries recognized immediately at the time of the initial surgery incurred a significantly lower cost and reduced hospital stay than those in whom recognition was delayed. As the surgical management of these injuries requires special multidisciplinary expertise, referral to and treatment in specialized centres offers the best chance of reversal from a potentially fatal condition to long-term restoration of good health with freedom from symptoms and return to normal liver function.

A long stricture of the lower end of the common bile duct, which characteristically forms an angle with the proximal dilated duct, is encountered in 15–20% of patients with chronic pancreatitis (Figure 25.64a). However, the differentiation between this disease and pancreatic carcinoma is often difficult in these patients. Bile duct strictures are also a feature of recurrent pyogenic cholangitis and sclerosing cholangitis. A similar long but less angulated stricture of the distal common bile duct may

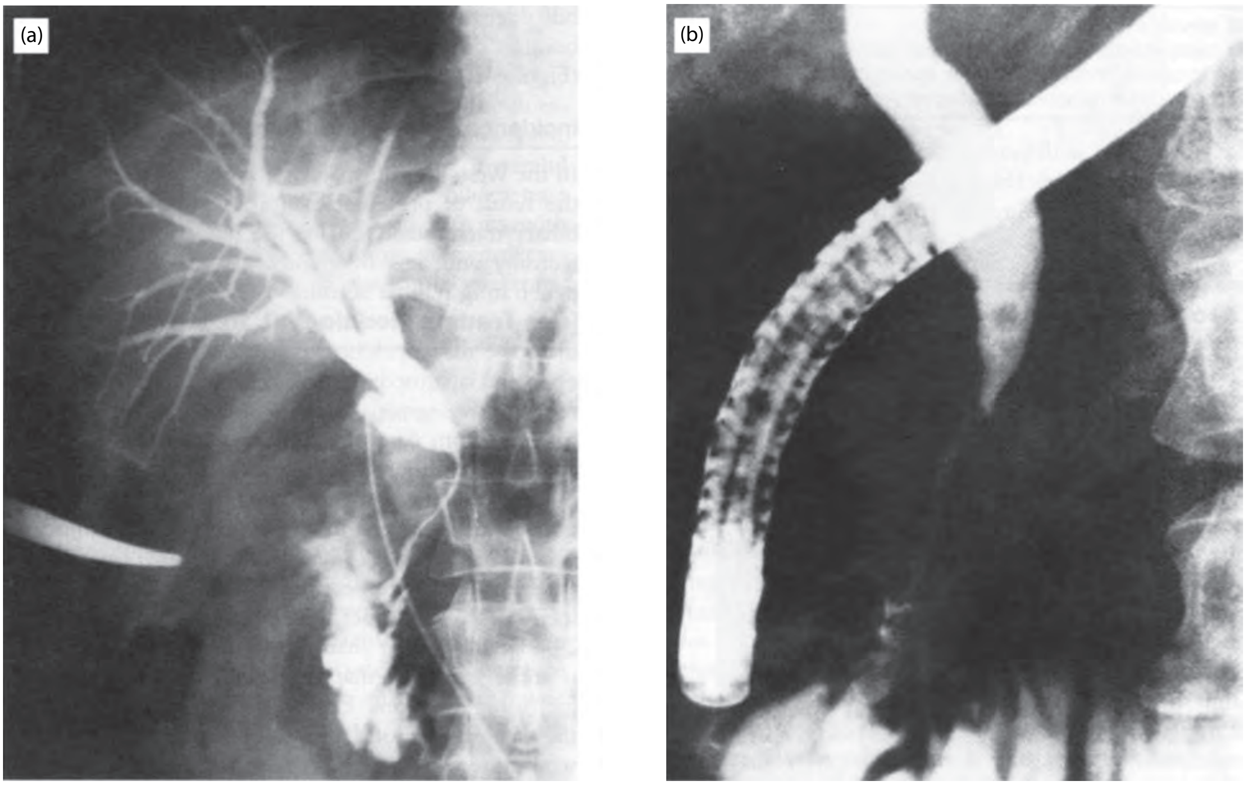


Figure 25.64 (a) Long stricture of the distal bile duct in a patient with chronic alcoholic pancreatitis of 12 years' duration. There is an angle between the stricture and the proximal dilated duct. The stenosis is due to the pancreatic fibrosis which constricts the transpancreatic segment of the common bile duct. (b) Stricture of the lower end of the common bile duct following blind stone fragmentation of an impacted stone and forcible passage of metal bougies.

be caused by trauma inflicted with metal bougies during open common bile duct exploration (Figure 25.64b).

Incidence of bile duct injury

In the West, the vast majority of bile duct strictures are the result of preventable injuries to the extrahepatic biliary tract, usually during the operation of cholecystectomy and, less frequently, gastrectomy. Blunt and penetrating trauma accounts only for a small minority

It is difficult to estimate the real incidence of bile duct injury after both open cholecystectomy and LC from published retrospective series. Accurate incidence can only be gathered from data of regional or national prospective audits that ensure complete data collection. Surveys from different countries reviewed by Strasberg *et al.* in 1995 showed an incidence of iatrogenic bile duct injury of 0.125% and 0.55% during open cholecystectomy and LC respectively. Based on this review, the incidence of bile duct injuries during sustained LC is 2–4.5 times higher than in the open technique. There is some evidence that the incidence decreases with increasing experience of the surgeons. This is suggested by the data of a national prospective audit in Switzerland, where the incidence of bile duct injury following LC fell from 0.6% to 0.3%. More recent reported series seem to confirm this positive trend and indicate that the incidence of biliary injury during open cholecystectomy and LC may be converging. The reason for this is probably twofold: increased experience with the laparoscopic technique and allocation of the more difficult cases to the open technique. In a recent population-based study of iatrogenic injury associated with cholecystectomy in Western Australia the risk of iatrogenic injury during LC was found to be 1.79 times that of open cholecystectomy.

Classification of bile duct injuries

The most commonly used classification of bile duct injuries is that reported by Corlette and Bismuth in 1981 based on an analysis of 643 cases of postoperative biliary strictures. The basis of this classification is the length of the proximal biliary stump, since this is the most important factor in determining the nature of the biliary repair.

- type 1 – low common hepatic stricture, length of common hepatic duct stump >2.0 cm (Figure 25.65)
- type 2 – middle stricture, length of hepatic duct stump <2.0 cm
- type 3 – high (hilar) stricture – no serviceable common hepatic duct but the confluence of the right and left hepatic ducts is preserved (Figure 25.66)
- type 4 – high stricture where the confluence is involved and there is no communication between the right and left hepatic ducts. The thickness of the fibrosis separating the two branches depends on the extent of the injury, i.e. thin or thick septum (1–2 cm) (Figure 25.67)
- type 5 – combined common hepatic and aberrant right hepatic duct injury separating both from the distal biliary tract.

The Corlette–Bismuth classification has proved useful because it provides essential information on the nature, risks and prognosis after the repair. There is an established correlation between the types of injury and the morbidity, mortality, success and recurrence after repair. However, the Corlette–Bismuth classification does not stipulate the length of the injury. This information is becoming increasingly important, as nowadays short strictures can be managed by non-operative treatment, such as percutaneous or endoscopic dilatation or stenting. A subclassification that indicates the extent of the lesion is desirable, i.e. discontinuity following excision of bile duct, short- or long-segment stenosis.

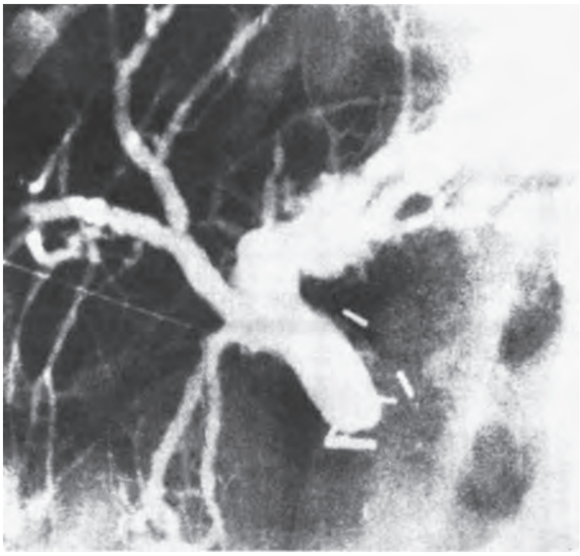


Figure 25.65 Type 1 Corlette–Bismuth injury.

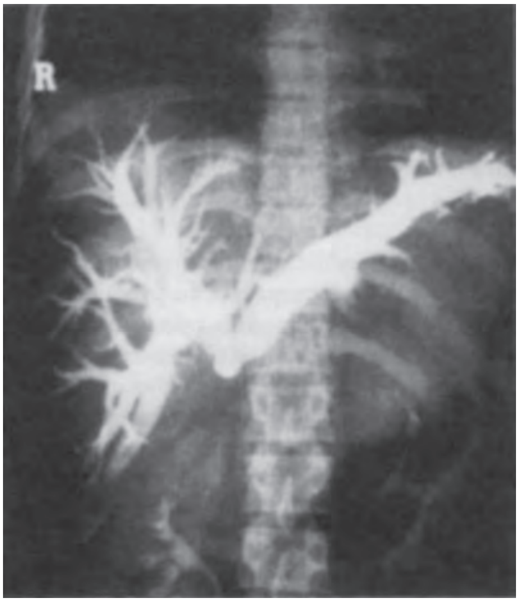


Figure 25.66 Obliteration of the hepatic duct with an intact confluence of the right and left hepatic ducts – type 3 stricture.

The more recent Strasberg classification considers bile injuries from a clinical perspective and includes biliary *complications* excluded in the Corlette–Bismuth types, e.g. bile leaks and bilomas, and isolated occlusion of the right hepatic duct. In essence it distinguishes two main categories:

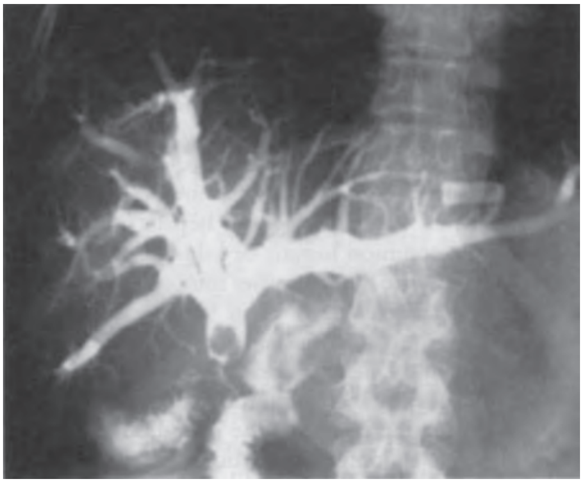


Figure 25.67 High stricture with destruction of the hilar confluence. A primary ductal stone straddles the two ducts – type 4 stricture.

- injuries that separate hepatic parenchyma from the biliary tract
- those where the continuity is maintained.

The classification also groups together injuries that have similar presentation and management, as distinct from those that require different management despite having similar presentations.

- *Type A* is a bile leak from a minor duct that is still in continuity with the common bile duct. These leaks occur either from the cystic duct stump or from the liver bed. These are not really bile duct injuries. Their importance lies in early recognition and appropriate management. They do not cause strictures or require tertiary referral.
- *Type B* is occlusion of part of the biliary tree – usually occlusion of an aberrant right hepatic duct mistaken for the cystic duct. In 2% of patients the cystic duct enters the right hepatic duct instead of the common hepatic duct. These injuries are often asymptomatic, or may present later with pain or cholangitis. The related hepatic segments drained by the occluded duct atrophy with variable hypertrophy of the rest of the hepatic parenchyma (disconnecting injury).
- *Type C* injury is bile leak from duct not in communication with the distal common bile duct. It is usually the result of transection of an aberrant right hepatic duct with drainage of bile into the peritoneal cavity. This lesion usually presents in the early postoperative period (disconnecting injury).
- *Type D* is lateral injury to the extrahepatic bile ducts. The hepatic parenchyma remains in communication with the distal end of the biliary tree and duodenum. Unlike type A, however, the consequences of type D injuries are potentially more serious. They require laparotomy for repair, and may result in stenosis. Type D injuries may involve the common bile, common hepatic, right and left hepatic ducts.
- *Type E* is circumferential injury of major extrahepatic bile ducts with separation of liver parenchyma from the lower ducts and duodenum (major disconnection injury).

The Strasberg classification is more suited to a multidisciplinary approach to the management of bile duct injury and for this reason is gaining wider acceptance.

Pathology of bile duct injuries

There is little doubt that the majority of bile duct injuries sustained during operation result from failure of appreciation of the precise anatomy of the area. The situations which predispose to or result in damage to the bile ducts at operation are varied and are usually but not always related to the degree of technical difficulty of the operation. Moossa *et al.* identified four mechanisms of bile duct injuries in open cholecystectomy, and later added a fifth one for LC:

- ligating or transecting the wrong duct
- occluding the lumen of the common bile duct during 'flush ligation' of the cystic duct
- compromise of the blood supply of the duct by excessive dissection
- trauma to the lumen of the common bile duct during exploration by manipulation or forceful 'dilatation'
- injury due to inappropriate application of energy source.

Although these are important, a more practical consideration of the mechanisms involved in bile duct injuries during LC identifies two basic error groups:

- misinterpretation of the anatomy
- technical errors.

Misinterpretation of the anatomy

All the available evidence indicates this as the dominant factor in the aetiology and it accounts for 70% of biliary duct injuries sustained during LC. This primary error is related to the fact that during LC the surgeon is operating on images of the operative field rather than reality. Visual psychological studies have shown that humans scan pictures by slow pursuit eye movements and, on the basis of the visual information relayed on the retina by this scanning process, the brain makes an interpretation. If the surgeon makes a snap interpretation, success or disaster depends on whether this is right or wrong. The vast majority of bile duct injuries during LC would be avoided if the surgeon questioned the initial identification of key structures to the point of absolute certainty. In one report based on review of operative videotapes of patients who sustained a duct injury, various misinterpretations were identified.

Technical errors

The most important technical error is hilar bleeding. This may result from failure to secure the cystic artery or damage to the right hepatic artery that is particularly at risk when it is looped towards the gallbladder neck. As frantic attempts are made to control bleeding, adjacent structures may be injured by electrocautery or clips. These injuries can be prevented if the surgeon has a strategic plan for this eventuality. The first step is to apply pressure on the bleeding area using adjacent tissues that can be easily grasped and laid over the bleeding area. Two instruments are then used: an atraumatic grasper and suction irrigation. When these are in the operative field, the compression is relaxed and the bleeding point identified by suction and grasped. If control is not achieved within a few minutes, conversion with application of the Pringle's manoeuvre is mandatory.

Other technical errors include excessive tenting of the cystic duct–common bile duct junction by too much traction on the gallbladder such that a clip used to secure the cystic duct grips the lateral wall of the bile duct resulting in a lateral injury/

partial occlusion. Bile leakage and bilomas in the absence of significant bile duct injuries are due to slippage of a clip on the cystic duct or too deep a plane of dissection on the liver bed during the detachment of the gallbladder.

Risk factors

These relate to surgeon, patient and local pathology.

Training and experience – the proficiency-gain curve

Studies in the pre- and postlaparoscopic era found this to be a more important risk factor than local operative factors such as inflammation or bleeding. The fact that iatrogenic injuries can still be inflicted by experienced surgeons if they ignore basic surgical principles has been documented by reports in both open cholecystectomy and LC. In laparoscopic surgery experience is important not only in respect of technical competence but also in terms of correct visual perception and interpretation of the image displayed on the monitor. The *proficiency–gain curve* (erroneously called learning curve in the surgical literature) appears to affect the risk of bile duct injury. For open cholecystectomy a large Swedish survey showed that most biliary injuries were made between the 25th and 100th operation per surgeon. The Southern Surgeons Club reported an initial high rate of bile duct injury (2.2%) during the first 13 laparoscopic cholecystectomies per surgeon. This rate fell to 0.1% for subsequent operations. Although experience is important in the reduction of bile duct injuries, it is clear that in the laparoscopic era there is no norm and the proficiency-gain curve is surgeon related. Thus bile duct injuries sustained during LC occur over a larger spectrum of experience by individual surgeons.

Improper use of energized dissection systems

With the deployment of energized systems, the important fact is that any thermal source (whether high-frequency electrocoagulation, laser or ultrasonic dissection) can cause damage to the portal structures if used incorrectly. The damage may not be apparent during the operation but can be severe and cause intractable strictures later. The perceived notion that ultrasonic dissection does not cause collateral proximity damage is not supported by experimental studies.

Patient factors

The problem of morbid obesity in the laparoscopic era varies considerably from patient to patient. Some present fewer problems than with open surgery, whereas others are less easy because of their internal fat deposition that obscures the anatomy of Calot's triangle. Fatty livers may be difficult to elevate and are easily lacerated. Although increased age and male gender are associated with increased postoperative mortality after cholecystectomy, they are not significant risk factors for major bile duct injuries.

Anomalous and morbid anatomy

This includes 'dangerous anatomy' and 'dangerous pathological conditions' predisposing to biliary injury. These are present in 15–35% of injuries.

Dangerous anatomy embraces aberrant (anomalous) anatomy, pathological conditions that obscure the view of vital structures such as adhesions, inflammatory phlegmon and excessive fat in the porta hepatis. Anomalies of the biliary tree are present in 10–15% of patients and are not usually identified preoperatively. The anomaly most likely to be involved in ductal injury is when an aberrant right hepatic duct inserts low into the common hepatic or bile duct, and is mistaken for the cystic duct. Another important anomaly is a short cystic duct. Variations in vascular supply also present dangerous situations, not only from the risk of compromise of the hepatic arterial supply (usually right hepatic) but also by increasing the likelihood of haemorrhage during the course of the operation. Adhesions from previous abdominal operations and pathological conditions such as inflammation can distort the anatomy and predispose to injury.

Dangerous biliary pathology includes chronic inflammation with dense scarring (the Mirizzi syndrome and the scleroatrophic gallbladder), fibrosis in the triangle of Calot and acute cholecystitis. A significant iatrogenic injury rate has been documented in patients undergoing LC for acute cholecystitis. Oozing of blood during the procedure with impaired visualization and anatomical distortion associated with the acute inflammation contribute to this increased risk. Nonetheless, the benefit of LC vs open cholecystectomy for acute cholecystitis has been confirmed by a prospective randomized clinical trial. This study showed a significant reduction in the postoperative morbidity in the LC arm (only 3% minor complications vs 23% major complications and 19% minor complications in the open cholecystectomy group). In adopting LC as the routine option, it must be stressed that the need for conversion is encountered in 20–25% of cases, and an early decision should be made to convert electively in the presence of obscured anatomy. Thus the valid approach to LC for acute cholecystitis is a flexible one. The procedure starts with an exploratory laparoscopy to assess the technical difficulty of the operation with particular reference to the structures in the triangle of Calot. A large distended gallbladder should be aspirated and lifted by a retractor rather than grasped. Large stones impacted in Hartmann's pouch that cannot be dislodged may prove problematic. The practical axiom is a simple one, i.e. if adequate exposure for a safe dissection cannot be obtained, or the gallbladder is gangrenous or perforated, the case should be converted. In some cases, a fundus first dissection of the gallbladder may be required. Dangerous biliary pathology also includes polycystic disease of the liver and portal hypertension caused by cirrhosis or schistosomiasis.

Prevention of bile duct injuries

Prevention of iatrogenic injuries to the bile ducts during LC relies on (1) thorough understanding of the anatomy, risk factors and the mechanisms of injury described, (2) image interpretative skills, (3) meticulous technique and (4) timely decision for elective conversion in the presence of difficult anatomy. There are no reliable preoperative indicators of the risks of biliary and vascular injuries during LC. Prevention of these complications, therefore, depends on the adoption of correct surgical technique and a low threshold for conversion. Since the major direct causes of biliary injury are misidentification of anatomy and technical errors, safety is entirely dependent on complete visualization, display and identification of the structures in the triangle of Calot. In this respect, the 30° laparoscope provides a better view of the anatomy, especially the common bile duct.

The vast majority of surgeons use clips to secure the medial end of the cystic duct, and only a minority ligate this duct. Bile leakage without bile duct damage is usually due to slippage of the clip. There are well-documented reports of internalization of these cystic duct clips into the bile duct where the clip acts as a nidus for stone formation several months later. These patients usually present with jaundice and cholangitis suggestive of bile duct stricture. The diagnosis becomes apparent on ERCP. These problems can be prevented by use of absorbable clips or ligation of the medial end of the cystic duct.

During the detachment of the gallbladder from its liver bed the dissection should be kept close to the gallbladder and above the fascial covering of the gallbladder bed. This avoids both bleeding from the hepatic parenchyma and injury to segmental ducts in segments IV/V of the liver. This is a much commoner cause of postoperative bile leakage than damage to a duct of Luschka, which is a very rare anomaly.

Differences between open and laparoscopic bile duct injuries

There are some important differences between the pathology of bile duct injuries sustained during open cholecystectomy versus LC. The laparoscopic injuries tend to be more extensive involving excision of a segment of the common bile duct, and extend to higher levels often involving proximal hepatic ducts (Figure 25.68). The majority (60–75%) are not immediately recognized during surgery. The average age of patients with laparoscopic bile duct injury is 43 years, compared with 56 years in open cholecystectomy. Instances of combined bile duct and vascular injuries have been reported after LC and they carry a poor prognosis. In one reported series four patients had simultaneous occlusion or extirpation of the right hepatic or common hepatic artery. In the immediate postoperative period, three of four patients with combined injuries had hepatic necrosis and/or abscesses with two patients requiring percutaneous or operative drainage. None of the remedial biliary anastomoses failed in the patients with isolated bile duct injuries. By contrast, two patients with combined injuries have had recurrent stenosis following reconstruction. Thus patients with major bile duct injuries should be evaluated for concomitant hepatic arterial injury as management and outcome are influenced by the absence of arterial blood flow to the injured bile ducts and to the liver.

Clinical features

Only one-third of bile duct injuries sustained during LC are detected at the primary operation with the majority being discovered at an average of 10 days postoperatively. If the lesion is not recognized at operation, the bile duct injury usually declares itself postoperatively by the development of pain, fever and external biliary fistula if a drain had been placed at operation. Evidence of peritoneal irritation with guarding and rebound is present due to leakage of bile in the peritoneal cavity, and subphrenic/subhepatic collections are detected on ultrasound. Overt signs of sepsis including positive blood cultures rapidly supervene and the patient develops jaundice, but this may not be severe or progressive in the presence of an external biliary fistula. Apart from LFTs and culture of the discharge and blood, visualization of the biliary tract by an ERCP to ascertain the presence of bile duct damage and its severity is mandatory and should be performed without any delay. MRCP is inferior to ERCP in this situation as, despite documenting the damage and collection of bile, it does not determine whether the leak is active. If the ERCP establishes the diagnosis, the patient should be referred to a tertiary hepatobiliary centre.

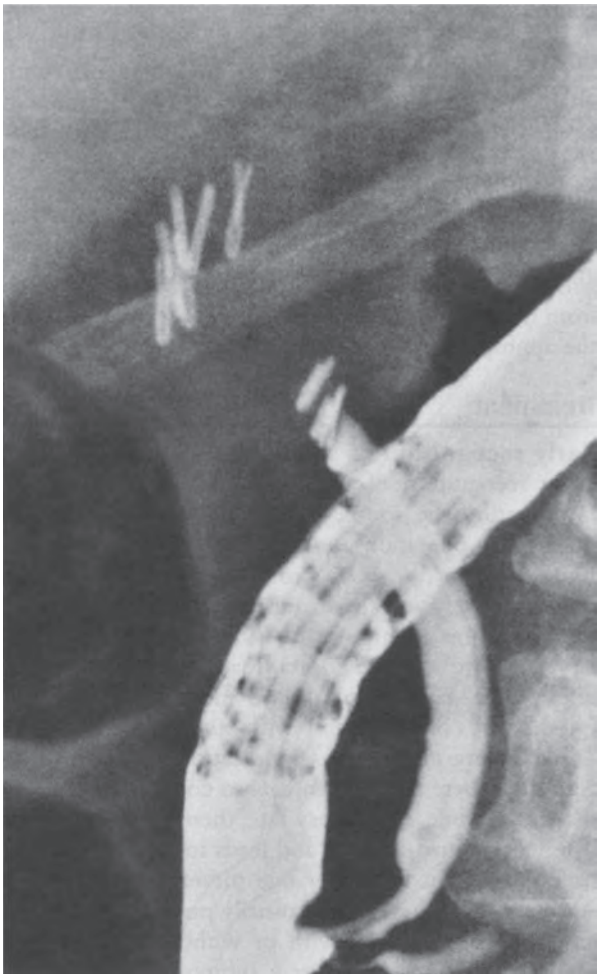


Figure 25.68 Bile duct injury after laparoscopic cholecystectomy: resection of the common hepatic duct. Note excessive shrapnel!

Some patients are actually discharged from hospital after LC and then admitted as an emergency with established sepsis and peritonitis. If a laparotomy is needed, the general surgeon should simply insert a self-retaining catheter into the proximal biliary tree through the defect in the bile duct, place a subhepatic drainage, evacuate any peritoneal bile collections and wash the peritoneal cavity. *Attempts at repair should be resisted at this stage.* As soon as the patient has recovered from the laparotomy, urgent referral to a tertiary unit is the appropriate course of action.

Treatment

Early recognition

Early recognition of the injury can be achieved by investigating the source of any biliary leak observed during the operation and routine IOFC. Only a few surgeons use IOFC routinely during LC, some only selectively and the majority not at all. Proponents of the routine use argue that IOFC delineates the biliary anatomy and provides a 'road map' of the entire biliary tree. Failure to outline the entire extra- and intrahepatic biliary tract with the patient in the Trendelenburg position is an indication for conversion. Although there are no firm data to confirm that IOFC reduces the bile duct injury rate, there is evidence that it reduces injury severity, and leads to detection at the time of surgery. Postoperative surveillance is crucial. Although bile duct injuries usually present postoperatively with a bile leak with or without jaundice and sepsis, some have only vague symptoms initially, e.g. persistent abdominal pain, anorexia, unwillingness to leave the bed. Abdominal pain persisting more than 12 hours after LC always requires investigation.

Definitive treatment

This depends on whether the injury is recognized at operation or subsequently because the patient either develops serious complications postoperatively or presents after discharge from hospital with recurrent episodes of pain, fever and jaundice.

Injuries recognized at operation

The treatment depends on the site and extent of the damage. For high complete transections, a Roux-en-Y hepaticojejunostomy is considered preferable to a difficult direct suture repair as this usually becomes strictured in time. For lower complete injuries with a serviceable proximal duct stump, primary suture repair with fine interrupted absorbable sutures over a T-tube is the treatment of choice (Figure 25.69). The long limb of the T-tube must not be exteriorized through the repair site as this enhances the risks of stricture formation. Partial (lateral) injuries are often treated by the insertion of a T-tube and a Roux-en-Y serosal patch. The long limb of the T-tube is exteriorized through the mobilized jejunum (Figure 25.70). Other techniques include repair with a vein patch over a T-tube (Figure 25.71).

There is an undoubted high incidence of stricture formation following primary repair of bile duct injuries (up to 60%). All these patients, therefore, require long-term follow-up with radiological assessment if they develop symptoms or abnormalities of the liver function tests.

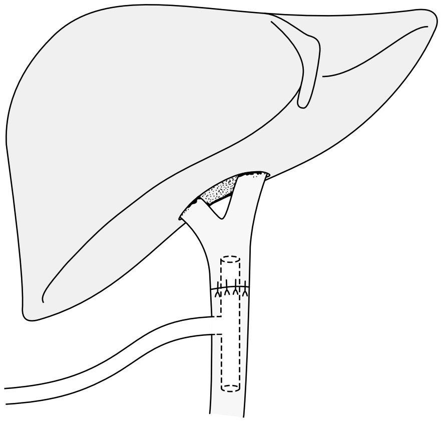


Figure 25.69 Primary repair of a complete injury recognized at operation. The long limb of the T-tube is exteriorized below the suture repair. The risk of stenosis is considerably enhanced if the long limb of the T-tube is brought out at the site of repair.

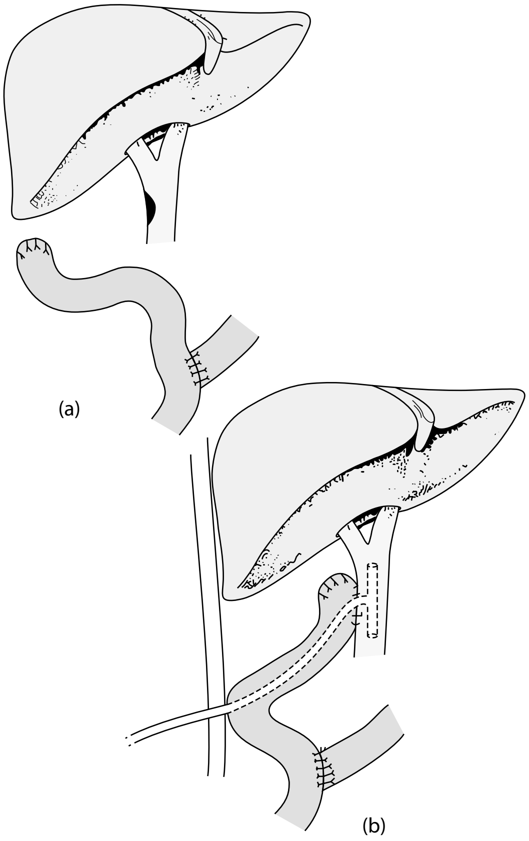


Figure 25.70 Serosal patch Roux-en-Y technique with T-tube intubation for lateral injuries discovered at operation. (a) Construction of a Roux-en-Y loop. (b) The Roux loop is sutured over the defect after the insertion of a T-tube, the long limb of which is brought out through the jejunal loop.

Injuries recognized in the postoperative period

The initial management is supportive. Fluid and electrolyte disorders, if present, are corrected and the patient is put on systemic antibiotics. Surgical intervention is required for drainage of collections/abscesses and development of peritonitis. Otherwise, the patient is initially managed conservatively. An ostomy bag is used to collect the bile leakage from the fistula, which may close. Persistence of the external biliary fistula does not constitute a serious problem since skin excoriation does not occur and the daily losses are seldom severe enough to cause significant fluid and electrolyte depletion. The bile collected in the bag is returned to the gastrointestinal tract via a nasoenteral tube, which is also used for feeding. Surgical intervention to deal with the bile duct injury at this stage is ill advised as the repair is difficult due to inflammatory oedema and because the proximal ductal system is not yet dilated. Repair is, therefore, best postponed for several weeks, by which time the intra-abdominal

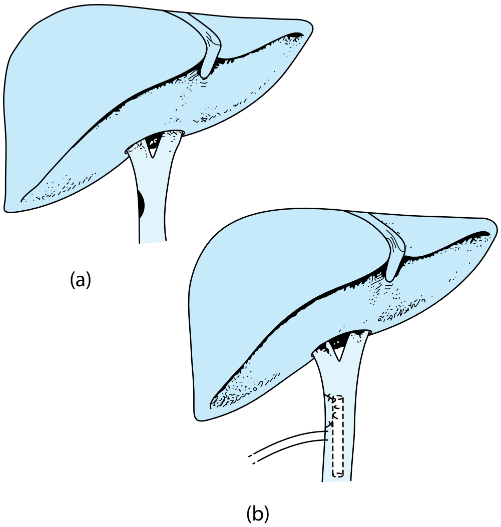


Figure 25.71 Repair of lateral injury recognized at operation by a vein patch. The T-tube is inserted via a small choledochotomy lower down.

sepsis has subsided, the stricture has matured and the proximal ducts have dilated, thus facilitating the procedure.

The definitive treatment of bile duct strictures is best carried out in specialized centres where the overall results, in terms of long-term freedom from jaundice and cholangitis and maintenance of good to normal liver function, are excellent (85–90%) and the operative mortality is low (1–5%). The exact treatment depends on the pathological anatomy of the stricture.

For type 1 injuries, the choice is between Roux-en-Y jejunostomy and choledochoduodenostomy (of the transection type). For common hepatic duct strictures with a serviceable extrahepatic duct stump (type 2), a Roux-en-Y hepaticojejunostomy is indicated (Figure 25.72). The establishment of a mucosa-to-mucosa anastomosis using absorbable fine sutures between the bile duct remnant and the jejunal mucosa is the single most important factor in the prevention of recurrent stricture formation. If good mucosal coaptation is achieved, there is no indication for stenting of the anastomosis and indeed this is undesirable under these circumstances. Some prefer to use an isolated jejunal isoperistaltic segment interposed between the stump of the common hepatic duct and the duodenum (Figure 25.73) instead of the Roux-en-Y loop.

For high strictures with no residual stump but with an intact hilar confluence (type 4), dissection of the liver plate (Glissonian sheath) to enter the hilum of the liver enables downward displacement of the hilar biliary confluence for a good mucosa-to-mucosa anastomosis between the bifurcation and the Roux-en-Y jejunal loop (Figure 25.74). The inside spur of the bifurcation is divided transversely and the edges are then sutured vertically to displace the junction of the two ducts proximal to the anastomosis. This technique does not require any splinting or resection of the liver parenchyma (Figure 25.75).

Destruction of the hilar confluence (type 5) constitutes the most difficult stricture to deal with and usually requires resection of the quadrate lobe to access the right and left hepatic ducts within the liver substance. Each duct is then anastomosed separately to the jejunal loop (Figure 25.76). In these high strictures, a silicone stent is placed across each anastomosis. These stents may be exteriorized transhepatically (Figure 25.77a) or preferably through the jejunal loop (Figure 25.77b) that is sutured to the abdominal wall at the exit site of the tube. The site of fixation of the jejunum is marked with metal clips. This allows easy identification of this access jejunostomy, thereby permitting the introduction of fine flexible endoscopes and balloon catheters for percutaneous dilatation in the event of restricturing after surgical repair. In the worst-case scenario when this technique is not technically possible because of dense fibrosis, the round ligament approach with anastomosis to the segment III duct (Figure 25.2) is used. This functions well

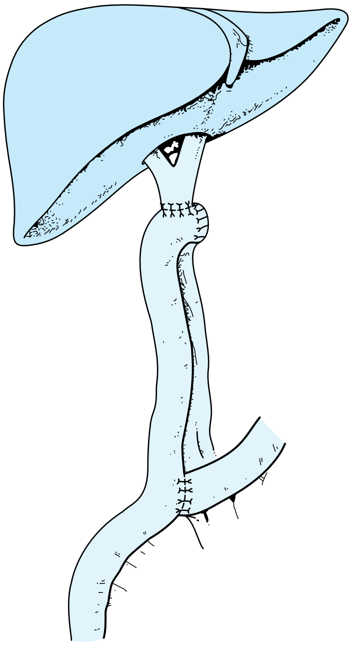


Figure 25.72 Roux-en-Y hepaticojejunostomy for a type 1 bile duct injury.

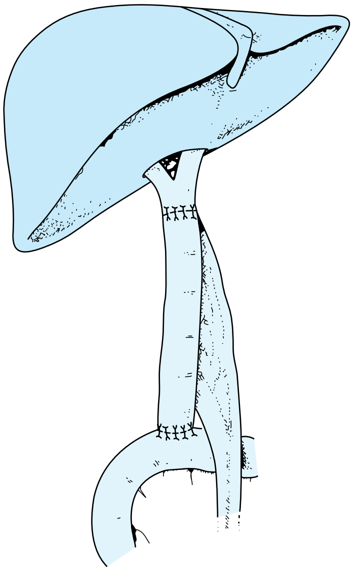


Figure 25.73 Isoperistaltic jejunal loop between the stump of the common hepatic duct and the duodenum.

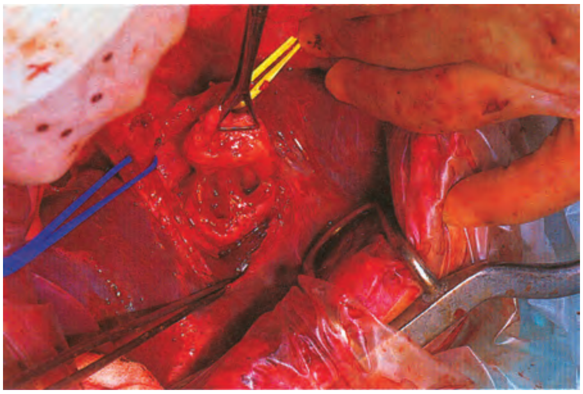


Figure 25.74 Operative photograph: repair of high stricture with no residual stump but with an intact hilar confluence (type 3). The liver plate (Glissonian sheath) has been divided anteriorly and posteriorly to enable entry into the hilum of the liver and downward displacement of the hilar biliary confluence for a good mucosa-to-mucosa anastomosis between the bifurcation and the Roux-en-Y jejunal loop.

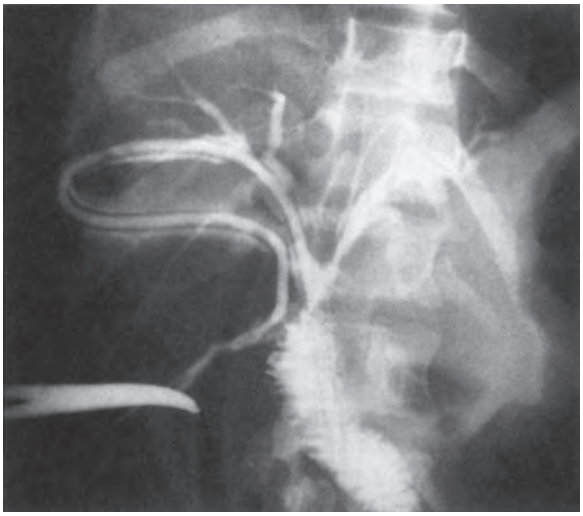


Figure 25.75 Postoperative cholangiogram through a transhepatic stent after Roux-en-Y hepaticojejunostomy for type 3 injury.

provided there is a communication between the right and the left ductal systems.

Endoscopic treatment of bile leaks, clipped ducts and benign strictures

Bile leaks

Bile leakage from slippage of the cystic duct ligature or clips (after LC or open cholecystectomy) is best treated by endoscopic means. Three methods are available:

- insertion of a stent which traverses the papilla and the origin of the cystic duct
- sphincterotomy to reduce the outflow resistance
- nasobiliary drainage with low negative pressure suction.

All appear to be equally effective. Endoscopic cannulation of the common duct has been used to dislodge clips partially occluding the common bile duct after LC. If successful, this is followed by stent insertion. It is not known at present how many of these patients will go on to develop bile duct strictures.

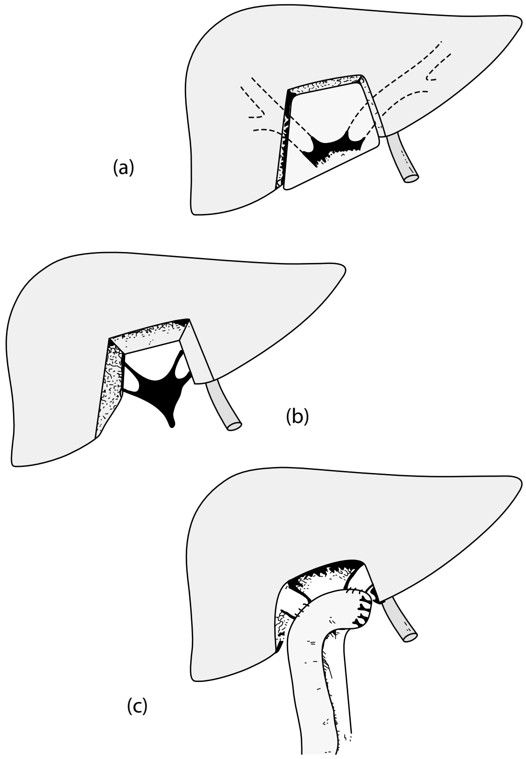


Figure 25.76 Hepaticojejunostomy for lesions with obliteration of the hilar confluence of the right and left hepatic ducts (type 4 injury): (a) the quadrate lobe (segment IV) anterior to the hilus is excised; (b) exposure of the obliterated confluence and identification of the right and left hepatic ducts; (c) the right and left hepatic ducts are anastomosed separately to the Roux-en-Y loop.

Bile duct strictures

Two techniques have been developed for the non-operative treatment of bile duct injuries:

- endoscopic balloon dilatation followed by stenting for 9–12 months
- PTBD using the Yamakawa prosthesis.

The ERCP technique is used whenever possible. A guide wire is passed through the stricture which is balloon dilated until the waist disappears, after which a plastic stent (10–11.5 Fr) is inserted. The stent is replaced every 3 months. Although metal stents have been used in the past, these have been abandoned because of two deaths associated with uncontrolled biliary sepsis. The reported long-term success of endoscopic stenting (stricture resolution) is 43% with a morbidity of 14% and mortality of 5%.

The PTBD is used when the ERCP technique is not possible. A special Yamakawa prosthesis is used. A peripheral bile duct is needled percutaneously under radiological control and a guide wire passed through the stricture into the duodenum (or jejunum following hepatojejunostomy). Initially a 10 Fr percutaneous drain is inserted over the guide wire. The drain is removed and the track dilated to 16 Fr some 7–10 days later when the Yamakawa prosthesis is inserted. The prosthesis is replaced every 3 months or sooner if it blocks and the treatment continued for

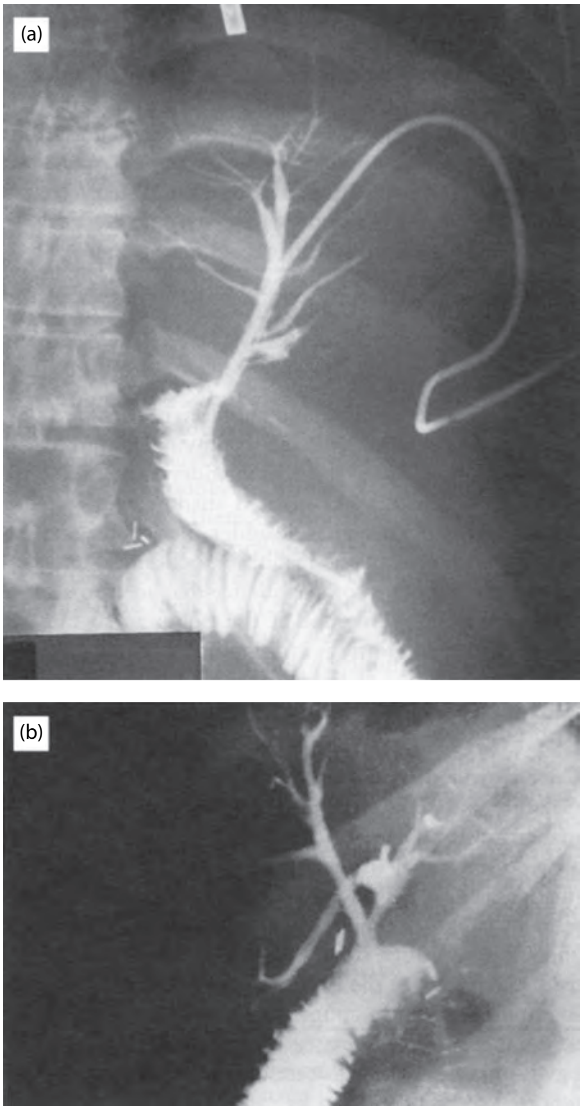


Figure 25.77 Postoperative cholangiograms after repair of type 4 injuries: (a) intrahepatic biliojejunal anastomosis has been stented transhepatically; (b) anastomosis has been stented through the jejunal loop.

12 months. The original Yamakawa prosthesis was manufactured from polyvinyl chloride. This material becomes brittle and breaks in bile. More recent versions do not have this problem. PTBD is more effective than the endoscopic technique is achieving stricture resolution but carries a higher morbidity (26%).

Obviously the results of non-operative treatment of bile duct strictures have steadily improved and there is now sufficient follow-up evidence to document stricture resolution by these methods. The treatment is however protracted and accompanied by a significant morbidity and a mortality of around 2–5%, i.e. similar to that of expert remedial surgery. The indications for non-operative treatment have not been clearly defined but recourse to endoscopic management seems sensible for:

- partial injuries
- poor-risk patients
- patients with portal hypertension
- failed surgical correction.

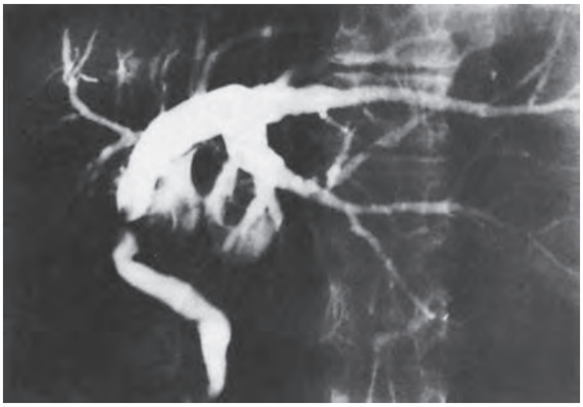


Figure 25.78 Localized stricture of the lower hepatic duct with ductal stones 10 years after cholecystectomy and exploration of the common bile duct. The patient had been completely symptom free until then and presented acutely with cholangitis. Some have ascribed these localized late strictures to impairment of the vascular supply.

Late bile duct injury

Rarely, there is no record of any untoward mishap during the operation and the patient has a smooth postoperative period and is discharged from hospital without symptoms or complications. Attacks of pain in the right hypochondrium with episodes of jaundice and cholangitis develop several months to years later. The majority of these cases are the result of unrecognized partial injuries sustained during the cholecystectomy. A vascular element due to damage to the ascending arteries of the bile duct has been postulated by some in the pathogenesis of some of the strictures. This is more likely to be the case in patients in whom a supraduodenal common bile duct exploration has been performed with T-tube drainage in addition to the cholecystectomy (Figure 25.78). Another cause of jaundice and cholangitis several months after cholecystectomy is internalization of metal clips used to secure the medial end of the cystic duct into the common bile duct. The clip acts as a nidus for the deposition of calcium bilirubinate. The biliary tract is infected in these patients. Not all 'strictures' of the bile duct presenting several years after a LC are the result of iatrogenic damage. The cholangiogram shown in Figure 25.79 of a patient who had undergone LC 5 years previously was misdiagnosed as a late stricture but turned out to be a CC at operation.

Tumours of the biliary tract

Tumours of the gallbladder

These consist of benign lesions and carcinoma of the gallbladder.

Benign tumours

Adenomas of the gallbladder form well-demarcated, polypoid lesions usually less than 2.0 cm in size. Most are discovered in clinical practice following pathological examination of the excised gallbladders, although some are identified as a fixed isolated shadow seen on the oral cholecystogram or gallbladder ultrasound scan in patients with unexplained right hypochondrial discomfort or

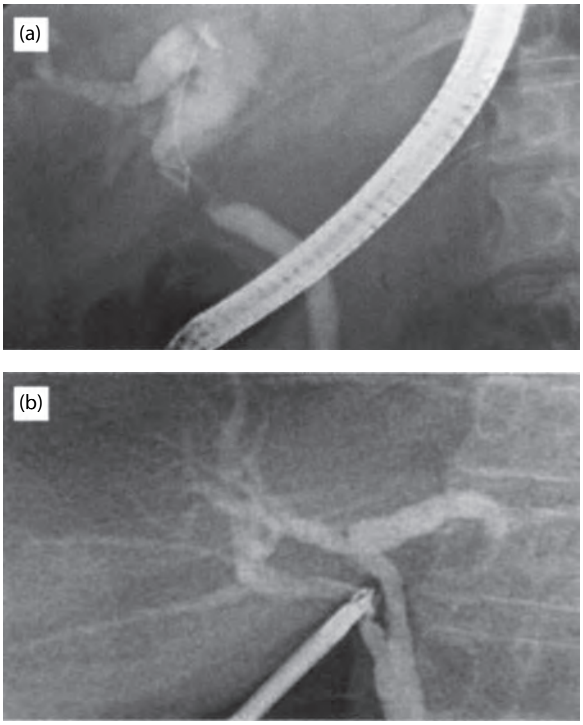


Figure 25.79 (a) Stricture of the common hepatic duct 5 years after laparoscopic cholecystectomy diagnosed preoperatively as iatrogenic injury with delayed presentation. (b) The operative cholangiogram at the time of surgery was normal. A cholangiocarcinoma was found and resected at operation.

pain. They are rare, being found in 0.3–0.5% of cholecystectomy specimens and are usually solitary. They are classified as tubular, papillary or tubulopapillary, depending on their growth pattern. There is some debate whether gallbladder adenomas may progress to carcinoma or not. On ultrasound they are typically echogenic when small, but become more heterogeneous as they enlarge. As ultrasound imaging alone does not identify any coexistent dysplasia or carcinoma *in situ* or indeed whether a polypoid mass in the gallbladder is benign or malignant, the current recommendation is for active intervention by cholecystectomy if the lesion is greater than 1.0 cm in size, in patients older than age 50 years for any polyp that is clearly growing in size on repeat ultrasound scanning even if less than 1 cm.

Carcinoma of the gallbladder

Gallbladder carcinoma represents 98% or more of all gallbladder malignancies; the rest comprise non-epithelial tumurs arising from the muscular or neurological components of the wall, metastases or lymphoma (lymphoma). Carcinoma of the gallbladder is the most common malignancy of the biliary tract and accounts for 3–4% of all gastrointestinal malignancies. The reported autopsy incidence is 0.6–1%. The disease is most commonly seen in elderly women (average age of 65 years) and affects females three times as commonly as males. Death from carcinoma of the gallbladder appears to be declining in many Western countries. This has been attributed to the increased cholecystectomy rate. The exact aetiology is unknown but gallstones are present in 75–90% of reported series of gallbladder cancer. Although the association with gallstones is well established, evidence that the stones are responsible directly for the cancer is not available. The role of infection and the consequent production of carcinogenic bile acid derivatives have been suggested but remain unconfirmed. There is undoubtedly an increased risk in patients with calcified gallbladder irrespective of whether this is patchy or diffuse (porcelain gallbladder). The increased risk is sufficiently great to warrant a cholecystectomy, even if the patient is asymptomatic. An increased incidence of gallbladder cancer is found in chronic typhoid carriers, South American Indians and obese individuals.

Pathology

The histological types of gallbladder cancer are:

- adenocarcinoma – vast majority, include papillary and undifferentiated
- squamous and adenosquamous – rare (5–7%)
- gallbladder neuroendocrine tumours and carcinoids – very rare (1%).

The *ras* family of genes consists of three functional genes that encode highly similar, guanine nucleotide-binding proteins (p21) with GTPase activity. The p21 protein is present on the inner aspect of the plasma membrane of a variety of cells. Strong ras p21 immunoreactivity has been documented in most gallbladder adenocarcinomas (62%) but not in gallbladder dysplasia or chronic cholecystitis. By contrast, ras p21 overexpression is encountered only in a minority of the cases of ampullary carcinoma and CCs. In gallbladder cancer, there is no significant correlation between ras p21 expression and patient survival or between ras p21 expression and p53 immunoreactivity. Furthermore there is no relationship to tumour grade. This suggests that ras p21 may be important in the development of gallbladder carcinoma but not in its progression. The lower rate of ras p21 overexpression in CCs and ampullary tumours indicates that these tumours probably have a different molecular origin.

The vast majority of gallbladder cancers are adenocarcinomas, with papillary, undifferentiated, squamous and adenosquamous constituting a minority. Rare tumours include carcinoid, melanoma and gallbladder neuroendocrine tumours (GB-NETs) which may secrete adrenocorticotrophic hormone.

There is limited published information on squamous cell and adenosquamous carcinomas of the gallbladder. In one recent large published series of 606 resected invasive gallbladder cancers, squamous differentiation was identified in 41 cases (7%). These were classified as pure squamous (without any identifiable glandular-type invasive component, $n = 8$) or as adenosquamous (tumours with the squamous component constituting 25–99%, $n = 26$). The incidence of adenosquamous carcinoma was 4%, and that of pure squamous cell carcinoma (without any documented invasive glandular component) was 1%. The clinicopathological characteristics of adenosquamous carcinoma/squamous cell carcinomas were contrasted with the more common gallbladder adenocarcinomas which constituted the vast majority. The average patient age was 65 years (range 26–81) with a female–male ratio of 3.8. Overall the diagnosis of gallbladder cancer was suspected peroperatively in a very small minority (13%). Macroscopically 58% presented with thickening and hardening of the wall and 6% were polypoid. Squamous metaplasia adjacent to the tumour was detected in 12%. All pure squamous cell carcinomas had prominent keratinization. Giant cells and tumour-infiltrating eosinophils were observed in 29% and 51% of the squamous cell carcinomas/adenosquamous carcinomas, respectively, vs 10% in gallbladder adenocarcinomas. All but three patients had 'advanced' (pT2 and above) disease. Follow-up of 31 patients showed that 25 died of disease (median survival of 5 months).

GB-NETs represent only 0.5% of all NETs. A recent review concluded that they originate from a multipotent stem cell or from neuroendocrine cells in intestinal/gastric metaplasia of the gallbladder mucosa secondary to cholelithiasis/chronic inflammation. GB-NETs are classified as typical and variant carcinoids. Clinically and at surgery, GB-NETs are indistinguishable from gallbladder cancer and are very rarely associated with the carcinoid syndrome (~1%). The median survival of GB-NETs is poor (10 months among 278 cases of GB-NETs reported in the Surveillance Epidemiology and End Results database). The 5 year survival rate for tumours classified as carcinoids/neuroendocrine carcinoma or small cell cancer is 36% and 0%, respectively. Thus GB-NETs have an aggressive behaviour, and, once diagnosed, extensive surgical treatment with careful follow-up with CT scan is mandatory to facilitate early detection of recurrence.

The Union of International Cancer Control (UICC) staging of gallbladder cancer is based on the depth of invasion as follows:

- stage I – confined to the mucosa/submucosa
- stage II – involvement of the muscle layer
- stage III – serosal involvement
- stage IV – spread to the cystic node
- stage V – invasion of the liver and adjacent organs.

Stage V is referred to as advanced gallbladder carcinoma and is classified into four types in accordance with the extent of involvement of adjacent organs:

- type I – hepatic involvement with or without gastrointestinal invasion (Ia, Ib)
- type II – bile duct involvement with or without gastrointestinal invasion (IIa, IIb)
- type III – hepatic and bile duct involvement with or without gastrointestinal invasion (IIIa, IIIb)
- type IV – major gastrointestinal involvement without hepatic or bile duct invasion.

Clinical features

The median age at presentation of gallbladder cancer is 65–70 years, with a female-to-male ratio of 3:1. The major risk factor is chronic cholecystitis, which is associated with dysplasia and carcinoma *in situ* and which subsequently progresses to invasive carcinoma. Other risk factors include (1) PSC, (2) anomalous junction of the pancreaticobiliary ducts and (3) choledochal cysts. Gallbladder carcinoma is found incidentally in 1% of cholecystectomy specimens and in up to 6% of cholecystectomies performed for polypoid lesions. It is usually not suspected clinically nor detected by preoperative imaging. Indeed when diagnosed by imaging, gallbladder cancer is usually advanced. The lack of serosa in the portion of the gallbladder attached to the hepatic parenchyma in essence means that the connective tissue on the undersurface of the gallbladder is part of the hilar plate and hepatic interlobar fissure, accounting for the early tumour invasion of the bloodstream, lymphatics, adjacent liver parenchyma and hepatoduodenal ligament.

The disease is either discovered accidentally during cholecystectomy or presents with non-specific symptoms or acutely with an inflammatory mass in the right hypochondrium (acute cholecystitis). The non-specific symptoms include upper right abdominal pain, which is the commonest symptom (76%), anorexia, nausea and vomiting (32%) and weight loss (39%). Jaundice is present in 38% at the time of presentation and is due to involvement of the common hepatic duct. The liver may be enlarged or the gallbladder may be palpable. Ascites is encountered in advanced disease. Anaemia is present in 50% of patients and is due to chronic haemobilia. Even in the presence of a normal serum bilirubin, the majority of patients have an elevated alkaline phosphatase activity. Few patients with stage I or II disease are diagnosed preoperatively.

Although ultrasound examination of the gallbladder readily identifies advanced disease, it misses the early potentially curable lesions. The most common ultrasound imaging presentation of the gallbladder is that of a mass completely replacing the gallbladder (40–65%), less commonly focal or diffuse wall thickening (20–30%) or an intraluminal polypoid mass (15–25%). Large gallbladder carcinomas which replace the entire gallbladder are often heterogeneous with obliteration of the normally distinct plane which separates the gallbladder from the adjacent liver. Other suggestive ultrasound findings include immobility of gallstones displaced by the mass. Hepatic arterial Doppler waveforms within an intraluminal mass are also suspicious for malignancy. Any solitary mass greater than 1.0 cm with internal vascularity should raise suspicion for gallbladder carcinoma. Although moderate wall thickening alone is not enough to distinguish gallbladder cancer from more common benign disorders, such as chronic cholecystitis or adenomyomatosis, marked wall thickening greater than 1.0 cm or loss of the normal mural layers of the wall should raise suspicion of gallbladder malignancy. Useful diagnostic information is obtained by angio-CT or MRI scan, which are used for tumour staging preoperatively. The highest diagnostic yield is obtained by laparoscopy with contact ultrasonography and this is nowadays regarded as the gold standard for the diagnosis and staging of this tumour.

Treatment

The treatment of cancer of the gallbladder is surgical though opinions vary as to the exact operative procedure that should be done. For stages I–III, the best results are reported with extended cholecystectomy. Initially, the gallbladder is removed and the diagnosis is confirmed by frozen section. If this is positive and the tumour is stage I–III, a 3.0 cm resection of surrounding hepatic parenchyma is performed together with radical lymph node clearance. Resection of the extrahepatic bile duct in continuity is performed if the tumour encroaches on the common hepatic duct.

If the tumour is advanced (stage V), some advocate an aggressive approach with right lobectomy or excision of segments IV and V together with extensive lymphadenectomy (cystic, hepatoduodenal, retroduodenal and coeliac nodes) and claim reasonable disease-free survival periods. Long-term survival has been reported after aggressive major resections of this nature only in patients with type I advanced gallbladder cancer. The other types of advanced disease, even if resectable (II–IV), are generally regarded as incurable and only palliative treatment is indicated in view of the uniformly poor prognosis.

A frequently encountered problem relates to patients in whom an unsuspected gallbladder cancer is found after LC on pathological examination of the resected specimen. Most would agree that the port wound through which the gallbladder was extracted should be excised full thickness. Thereafter, there are no firm guidelines, but it would seem sensible that if the cancer is pT1 (confined to mucosa/submucosa) then no further action is needed except for careful follow-up of the patient. If the tumour involves the muscularis or serosa, then hepatic resection (segments IV and V) and lymph node clearance is wise provided the patient's general condition is good.

In inoperable patients with jaundice and itching, palliation can be achieved either by endoscopic/radiological stenting with metal expanding endoprostheses or surgically by the round ligament segment III bypass. The palliation produced by this simple surgical bypass that is well tolerated is very good. The response of gallbladder cancer to radiotherapy and chemotherapy is poor.

Survival

The overall 5-year survival of patients with gallbladder cancer is 4%. The reported cumulative 5 year survival after extended cholecystectomy is 90–100% for stage I and 50–65% for stage II disease. Long-term survivors may also be encountered in patients with stage III, node-negative tumours (T3N0). The survival of patients with squamous cell carcinomas/adenosquamous carcinomas is significantly worse than that of gallbladder adenocarcinomas, even when compared with stage-matched advanced gallbladder adenocarcinoma cases.

Tumours of the bile ducts

Benign tumours

A variety of benign tumours of the bile duct including adenoma and papilloma have been reported but they are rare and far less common than CCs. Benign bile duct tumours have a tendency to recur after excision and some have been reported to undergo malignant change. Benign bile duct tumours present with jaundice and occult chronic gastrointestinal haemorrhage (haemobilia).

Malignant tumours

The reported autopsy incidence of malignant bile duct tumours ranges from 0.01% to 0.5%.

Malignant bile duct tumours are by tradition in two groups: (1) CCs and (2) periampullary tumours.

Cholangiocarcinoma

CC arises from the epithelium of the intrahepatic and extrahepatic ducts excluding the gallbladder and the papilla of Vater. The annual age-standardized incidence in Europe is 1.0–1.5 cases per 100 000 population although in both Europe and the USA the incidence has been gradually rising over the last three decades. CC is more common in males (in contrast to gallbladder cancer) and this is largely attributed to the higher incidence of PSC in men since this disorder predisposes to CC. There is a marked geographical variability in the incidence of CC attributed to regional environmental risk factors. The highest incidence of CC is in Thailand closely followed by Japan. In Europe, the highest incidence is in Denmark. The majority of patients are old (>65 years) with the peak incidence occurring in the seventh decade. The prognosis is generally poor with a 5 year survival of less than 5%. Most patients present with advanced disease which precludes surgical resection. These patients die within 1 year of diagnosis from cachexia, liver failure and recurrent sepsis due to biliary obstruction. The reported 5-year recurrence rate after surgical resection varies from 60% to 90% although the 1 year survival has increased from 16% in 1975–1979 to 28% in 1995–1999. However, as expected the 5-year survival has not.

Aetiology

Most CCs arise sporadically and the exact cause remains unknown. However, there is good evidence that chronic inflammation and biliary duct cell injury induced by the obstruction are responsible for the development of CC. Cytokines released by the inflammation are held responsible for the malignant transformation. In some 10% of patients there are established risk factors/disorders, all of which cause a chronic inflammation of the bile ducts. In Western countries the most common disease associated with CC is PSC. One large Swedish study documented a CC incidence of 8% in patients with PSC during a mean follow-up period of 5 years. Seventy per cent of these patients also suffered with ulcerative colitis. Occult CC is reported in up to 36% of autopsies and in 40% of livers removed in patients undergoing hepatic transplantation for PSC. CC in PSC patients tends to occur at a younger age (30–50 years), is frequently multifocal and is usually not resectable. Another risk factor is fibropolycystic malformation of the biliary tree, e.g. choledochal cysts, and especially Caroli disease. As in PSC, the CC presents at an earlier age than the sporadic form. The increased risk of tumour is thought to be related to an abnormal choledochopancreatic duct junction, which predisposes to reflux of pancreatic secretions into the biliary tree, chronic inflammation and bacterial contamination. Parasitic infestation of the liver prevalent in Japan and Southeast Asia (*Opisthorchis viverrini, C. sinensis*) also predisposes to CC, which develops in 8–10% of cases. Infestation occurs from consumption of raw or undercooked fish containing adult worms, which then lay their eggs in the biliary system. The parasite induces a chronic inflammation in the biliary tree, which increases the risk of CC. Infestation with *O. viverrini* is a major public health problem in north-east Thailand, where approximately one-third of the population is infected. CC is one of the leading causes of death in this region. The median age at presentation of CC in areas of endemic infestation with *O. viverrini* is 52 years (range 32–69) and the majority of patients are male. Population-based studies using ultrasonography to visualize early tumours have documented the close association between CC and heavy infestation with this parasite. Survival after surgical treatment of CC in patients with opisthorchiasis is broadly similar to that reported for CC without liver fluke infestation. There is also evidence that CC is associated with cirrhosis of any cause. Patients testing positive for hepatitis C virus or hepatitis B virus surface antigen have an increased risk of developing intrahepatic CC (ICC). Exposure to chemical agents and radionuclides has been identified as a risk factor for CC (nitrosamines, dioxin, asbestos, Thorotrast, radon). There is no firm evidence that excess alcohol consumption predisposes to the disease. Recently obesity was identified as a risk factor for the development of extrahepatic CC.

The association with gallstones is much less marked than it is with carcinoma of the gallbladder, but ductal calculi are found in 20–50% of patients who develop CC. Bacterial induced endogenous carcinogens derived from bile salts (e.g. lithocholate) have been implicated and their role is supported by the findings of some epidemiological studies and the higher incidence in typhoid carriers. Thus aside from sporadic cases, CC is seen with increasing frequency in certain clinical disorders, which include:

- parasitic infestation of the biliary tract
- PSC
- ulcerative colitis
- cystic disease of the biliary tract, especially Caroli disease
- chronic typhoid carriers
- ulcerative colitis.

Molecular oncology of cholangiocarcinoma

Several mutations in oncogenes and tumour suppressor genes have been identified in CC, and these are thought to induce and drive the pathogenesis of CC through a stepwise transformation of normal epithelium into carcinoma *in situ* and invasive CC. The mutations which have been identified include the genes k-*ras*, c-*myc*, c-*neu*, c-*erbB2*, c-*met*, *p53* and *bcl*-2. The overexpression of the *p53* tumour suppressor gene or mutations of the k-*ras* oncogene are associated with increasing cytological atypia and tumour aggressiveness. The increased c-*met* expression is thought to play a role in promoting metastases. Experimentally, overexpression of Bcl-2 reduces apoptosis in CC cell lines, while defective Fas receptor or increasing Fas ligand expression enable the tumour to escape immune surveillance. Two markers (fascin and mesothelin) are expressed with transition from carcinoma *in situ* to invasive adenocarcinoma, implicating a role of these markers in neoplastic progression. The majority of CCs (85%) express *p53* to a varying extent. By contrast the bile duct epithelium adjacent to the tumour and dysplastic areas do not. Thus overexpression of *p53* is thought to play an important role as a late event in the pathogenesis of these tumours. Marked diffuse p53 staining encountered in 30–40% of CC is associated with a significantly poorer survival than tumours with low/focal positivity. This adverse effect of high *p53* expression on survival is reported to be independent of age, sex, tumour size, radicality of resection, histopathological grading, lymph node status, perineural invasion and vasoinvasive growth.

Pathology of cholangiocarcinoma

The tumours are best classified into the anatomical site of origin:

- ICC – from the minor hepatic ducts (10%)
- hilar CC – from the right and left hepatic ducts, hilar confluence and proximal common hepatic duct (Klatskin tumours) (50–60%) (Figure 25.80)
- middle CC from the distal common hepatic duct, cystic duct and its confluence with the common bile duct (Figure 25.81)
- distal CC from the distal common bile duct (Figure 25.82); note that middle and distal are often grouped together as distal.

The classification of ICCs and their treatment remains controversial. Some are indeed diffuse (multicentric) and difficult to differentiate from primary hepatocellular carcinomas with which they are grouped because of similar clinical course and poor prognosis. However, the majority of ICCs fall into two macroscopic forms: *mass-forming type* (the majority) and the *periductal infiltrating type.* ICCs of the periductal infiltrating type have a marked tendency to spread along the nerves and lymphatics of the Glissonian sheath, whereas ICCs of the mass-forming type tend to invade the hepatic parenchyma via the portal vein, and exhibit perineural and lymphatic invasion of

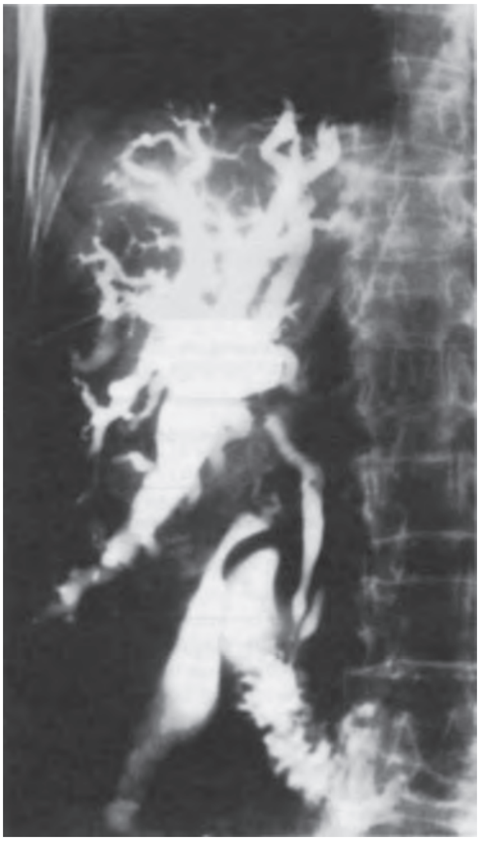


Figure 25.80 Percutaneous transhepatic cholangiogram in a 75-year-old female who presented with marked cholestatic jaundice showing a hilar (Klatskin) tumour.

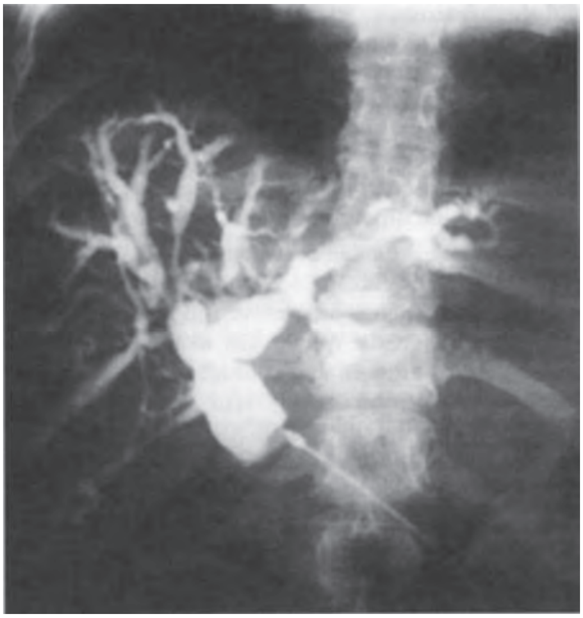


Figure 25.81 Operative cholangiogram showing a tumour at the junction of the cystic duct with the common bile duct. The patient presented with jaundice and acute cholecystitis.

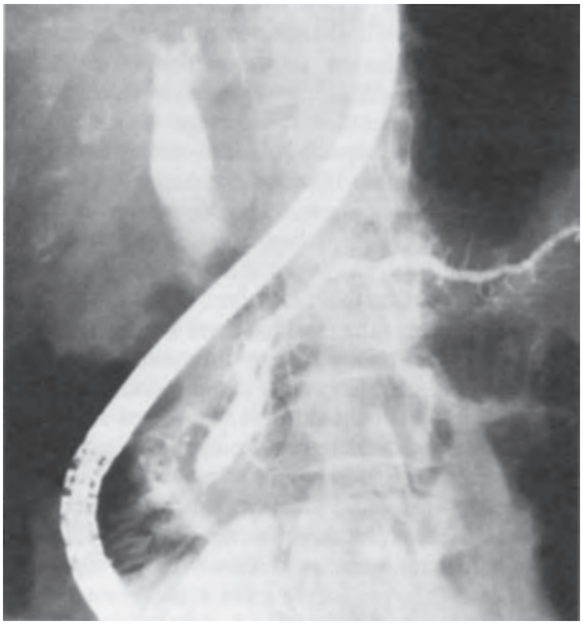


Figure 25.82 Endoscopic retrograde cholangiopancreatography showing a papillary tumour of the distal bile duct extending to the periampullary region.

the Glissonian sheath only when large. Both types can produce satellite hepatic deposits. Therefore, major hepatectomy with combined resection of the extrahepatic bile duct is recommended for all ICCs of the periductal infiltrating type and for those of the mass-forming type with invasion of the Glissonian sheath.

The gross appearances of extrahepatic CC (hilar and distal) assume one of three forms: stricture (scirrhous variety); nodular; or papillary. The scirrhous variety can be very difficult to distinguish from sclerosing cholangitis even on histological grounds. These tumours are generally confined to the proximal ducts (hilar) and form grey annular thickenings with clearly defined edges. The nodular tumours form extraductal nodules in addition to intraluminal projections. The papillary variety is most commonly found in the distal bile duct. These lesions are friable and may fill the duct lumen with vascular neoplastic tissue and tend to bleed in the ductal lumen causing haemobilia. The majority of tumours are adenocarcinomas of varying differentiation. The scirrhous variety is intensely fibrotic and relatively acellular, often with a few well-differentiated ductal carcinoma cells grouped as acini in a dense connective tissue stroma. Rare types include squamous cell carcinoma, adenosquamous carcinoma, adenoacanthoma, lymphoma, carcinoid tumours and melanoma. Malignant smooth muscle tumours of the bile duct have also been reported, as have instances of primary non-secreting apudoma of the hilar region. All CCs are slow growing, locally infiltrative and metastasize late. CCs have a special predilection for perineural spread and do not metastasize beyond the liver.

The commonest type is hilar CC (Klatskin tumours). These have been classified by Bismuth and Corlette into four types which are important in the surgical management:

- type I – tumour involves the common hepatic duct
- type II – tumour involves the bifurcation of the common hepatic duct
- type IIIa – tumour involves the right hepatic duct
- type IIIb – tumour involves the left hepatic duct
- type IV – tumour involves both the right and left hepatic duct.

The special pathological features of CC can be summarized as follows:

- early invasion of adjacent organs (when the tumour is located to the hilus, direct invasion of the liver or perihepatic structures such as the hepatic artery and portal vein)
- invasive spread with neural, perineural and lymphatic involvement, and subepithelial extension
- limited tendency to metastasize, and only one-third of patients demonstrate lymph node, hepatic or peritoneal metastasis at the time of surgery.

Over 90% of CCs are well-differentiated and mucin-producing adenocarcinomas. The macroscopic subtypes of CC are (1) sclerosing tumours, (2) nodular, (3) mixed sclerosing and nodular and (4) papillary. The papillary type is a low-grade adenocarcinoma associated with a better prognosis. Survival is related to the pathological grade of the tumour. The prognosis is worse in younger patients (<40 years). The various histological types (UICC) are:

- UICC – CC (extrahepatic)
  - carcinoma *in situ*
  - adenocarcinoma, NOS
  - adenocarcinoma, intestinal type
  - mucinous adenocarcinoma
  - clear cell adenocarcinoma
  - signet-ring cell adenocarcinoma
  - adenosquamous carcinoma
  - squamous cell carcinoma

  - small cell (oat cell) carcinoma
  - undifferentiated carcinoma (spindle and giant cell type; small cell type)
  - papillomatosis
  - papillary carcinoma, non-invasive
  - papillary carcinoma, invasive
  - carcinoma, NOS
- malignant mesenchymal tumours (rare)
  - embryonal rhabdomyosarcoma
  - leiomyosarcoma
  - malignant fibrous histiocytoma.

Classification of extrahepatic cholangiocarcinoma

This is based on the UICC–TMN (2002) classification.

Primary tumour (T)

- Tx: primary tumour cannot be assessed
- T0: no evidence of primary tumour
- Tis: carcinoma *in situ*
- T1: tumour confined to the bile duct histologically
- T2: tumour invades beyond the wall of the bile duct
- T3: tumour invades the liver, gallbladder, pancreas and/or unilateral branches of the portal vein (right or left) or hepatic artery (right or left)
- T4: tumour invades any of the following: main portal vein or its branches bilaterally, common hepatic artery, or other adjacent structures, such as the colon, stomach, duodenum, or abdominal wall

Regional lymph nodes (N)

- Nx: regional nodes cannot be assessed
- N0: no regional lymph node metastasis
- N1: regional lymph node metastases (at least three lymph nodes)

Distant metastasis (M)

- Mx: presence of distant metastasis cannot be assessed
- M0: no distant metastasis
- M1: distant metastasis

Stage grouping of extrahepatic cholangiocarcinoma

- Stage 0 (carcinoma *in situ*): TisN0M0
- Stage IA: T1N0M0
- Stage IB: T2N0M0
- Stage IIA: T3N0M0
- Stage IIB: T1N1M0; T2N1M0; T3N1M0
- Stage III: T4 any NM0
- Stage IV: any T any NM1

TNM classification of ICC (UICC 2002)

Primary tumour T

- Tx: primary tumour cannot be assessed
- T0: no evidence of primary tumour
- T1: solitary tumour without vascular invasion
- T2: solitary tumour with vascular invasion or multiple tumours but none >5 cm
- T3: multiple tumours >5 cm or tumour invading major branch of portal or hepatic vein(s)
- T4: tumour(s) with direct invasion of adjacent organs other than gallbladder or perforation of the visceral peritoneum

Regional lymph nodes (N)

- Nx: regional nodes cannot be assessed
- N0: no regional lymph node metastases
- N1: lymph node metastases in hepatoduodenal ligament (at least three)

Distant metastases (M)

- Mx: presence of distant metastases cannot be assessed
- M0: no distant metastases
- M1: distant metastases

Stage grouping of ICC

- Stage I: T1N0M0
- Stage II: T2N0M0
- Stage IIIa: T3N0M0
- Stage IIIb: T4 N0M0
- Stage IV: any T any N M1

Prognostic factors

Prognosis is not influenced by the location of the cancer if radical surgery is performed. After surgical resection of ICC, 5-year survival rates vary from 8% to 47%. The best prognosis has been observed in patients with negative resection margins. For extrahepatic CC, 5 year survival rates of 20–54% have been reported. Only histological margin status and lymph node involvement are the important prognostic factors influencing survival. In ICC recent data suggest that poor prognosis is related to angiogenesis and intrahepatic metastasis. Survival is also related to the macroscopic subtype (better for papillary) and grading of tumour (better if well differentiated). Some authors also suggest that younger patients (<40 years) with peripheral CC have a significantly worse survival rate than older patients.

Clinical features

These reflect the location of the tumour. Painless jaundice is the leading symptom (>90%) in patients with extrahepatic CC. It is less frequent in patients with ICC. The biliary obstruction is accompanied by pale stools, dark urine and pruritus. Physical signs include hepatomegaly, right upper abdominal pain, weight loss and fever. Weight loss is not evident until the disease is advanced and is usually accompanied by evidence of hepatic involvement and ascites. Dull upper abdominal pain is a frequent symptom. Some patients present acutely with cholangitis or acute cholecystitis. The duration of symptoms is usually short and measured in months. Physical examination reveals hepatomegaly. Anaemia is present in patients with papillary tumours. It is caused by chronic haemobilia. The faeces of these patients have a characteristic silvery appearance due to a combination of steatorrhoea and altered blood. A significant percentage of patients with hilar tumours have previously undergone recent cholecystectomy (within 6–12 months of diagnosis). Regrettably, these tumours are missed at operation since the small nodule in the porta hepatis is not easily palpable, the common bile duct is not dilated and there is free flow of contrast into the duodenum. Often the surgeon concerned ignores the fact that there is poor filling of the intrahepatic biliary tree or interprets a localized narrowing as extrinsic vascular compression. An operative cholangiogram should never be passed as normal unless there is adequate and complete filling of the intrahepatic biliary tree.

In the preoperative assessment the important disorders which require exclusion are cancer of the pancreatic head, duodenal, gallbladder, periampullary cancers, ductal calculi, PSC and Mirizzi syndrome. PSC can pose problems as it is one of the important predisposing disorders. Aside from full LFTs which confirm cholestasis, the serum level of the carbohydrate antigen 19-9 (CA19-9) is a useful marker as its elevation >100 U/mL has a sensitivity of 89% and a specificity of 86% for CC. Serum CA19-9 is thus useful in the diagnosis. It is also used after radical resection and in monitoring the effect of treatment. After curative resection, serum levels decrease from a preoperative value although this is not a universally reported finding.

The initial imaging tests used most commonly are ultrasound or CT scan, which document the level of biliary obstruction, dilatation of the biliary tract and the absence of stones. The sensitivity of ultrasound in demonstrating the site of biliary obstruction has improved over the years and now exceeds 90%. Indeed duplex ultrasound is now considered equivalent to CT scan, portography and angiography in predicting vascular invasion and resectability. A helical CT scan can detect CC greater than 1.0 cm, the relationship between tumour and adjacent organs and vascular structures. It predicts accurately resectability of the tumour in 60%. MRCP has a similar sensitivity and specificity to ERCP. MRI is currently regarded as the best imaging modality for diagnosis and staging of CC (Figure 25.83). In specialist centres, ^{18}F-fludeoxyglucose positron emission tomography is used to detect small CCs and distant metastases and appears to be very useful in detecting CC in patients with PSC. Cytological diagnosis is obtained by ERCP or PTC. However, both techniques are largely restricted to therapeutic biliary drainage when indicated. There is now good evidence that only laparoscopic exploration can define the true resectability.

In patients with ICCs, mild abdominal pain is the most frequent presenting clinical sign and always warrants investigation. Presentation with jaundice is uncommon (13%). If a mass lesion is detected by imaging (CT, MRI, etc.), preoperative biopsy is ill advised since it may result in local implantation and in one reported series was correct in only 40%. In high prevalence areas such as Taiwan, the preoperative diagnosis of ICCs is difficult because of the frequent association with hepatolithiasis (43%) and the high incidence of hepatocellular carcinoma. Thus in one large reported series of 125 patients from Taiwan, 25% of these patients underwent surgery for chronic cholangitis and 12.5% for hepatocellular carcinoma rather than CC.

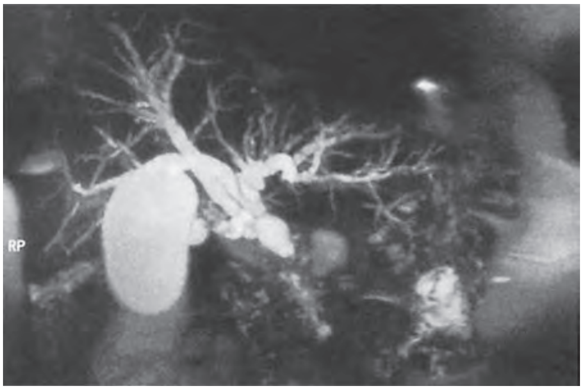


Figure 25.83 MRI/magnetic resonance cholangiopancreatography showing an advanced cholangiocarcinoma with involvement/thrombosis of the portal vein. The patient was treated by stenting.

Radiological signs of non-resectable extrahepatic CC include (1) bilateral hepatic duct involvement up to secondary radicals, (2) encasement or occlusion of the portal vein proximal to its bifurcation, (3) atrophy of one liver lobe with contralateral portal vein branch encasement, (4) involvement of both hepatic arteries, atrophy of one liver lobe with contralateral bile duct involvement and (5) distant metastases.

Preoperative tissue diagnosis of CC can be difficult because of its location, size and desmoplastic fibrous nature of many lesions. Bile cytology can be obtained with fine-needle aspiration with ultrasound or CT guidance; brush cytology can also be procured during ERCP, or by endoscopic transpapillary biopsy or PTC.

In the West, distal bile duct stones and proximal extrahepatic malignant biliary obstructions may coexist. These stones probably predate the development of the malignant obstruction and are found in 12–18% of patients with proximal tumours. In most instances the stones are distal to the malignant obstruction. They may interfere with stent function in inoperable cases.

The proximal (hepatic) longitudinal extension of hilar CC is difficult to predict before surgery. In this respect, the cholangiographic findings may be misleading. In one reported study of 54 patients, histological examination of specimens indicated proximal longitudinal spread of the tumour in 41%. Cholangiography was unreliable in detecting this spread (63% accuracy). A significantly higher frequency of longitudinal spread was observed when the cholangiographic images showed a main tumour with collapsed edges as distinct from tumours with sharp edges on the cholangiographic films. The authors recommend a mapping biopsy performed by percutaneous transhepatic cholangioscopy to establish the limit of the proximal spread but this is not practised in other centres.

Treatment

Surgical treatment

Surgery provides the only means of cure if a potentially curative resection is performed (complete excision of tumour with negative margins). However, the radical resection rate varies with the location of the CC: up to 70% in distal cancers as distinct from 15–20% in high CC. Radical resection is indicated if:

- the patient's general condition and nutritional status are good based on clinical assessment and laboratory tests (serum bilirubin and normal albumin levels); fit but overtly jaundiced patients should have preoperative biliary drainage
- imaging tests exclude extensive tumour, vascular invasion, hepatic lobar atrophy, metastatic disease. Portal vein involvement is always an independent predictor of unresectability.

However, preoperative assessment of resectability is not always possible, and in some cases the resectability can only be determined at operation. Either preprocedural laparoscopy or a limited laparotomy via a right subcostal incision can be performed for adequate staging in these patients. Management of perihilar CC depends on the structures involved and the degree of right or left hepatic duct or common bile duct involvement. The goals of surgical resection are complete excision of tumour with negative margins with restoration of bilioenteric continuity. Aggressive surgery (bile duct excision and partial hepatectomy with *en bloc* caudate lobectomy) is indicated for microscopically negative margins (R0 resection). Some surgeons advocate a more limited hepatic resection for centrally located tumour: resections which include segments IV and V or IV and I. Hilar CC is deemed unresectable in the presence of bilateral portal venous involvement or bilobar extensive hepatic involvement. Middle/distal CC is treated with pancreaticoduodenectomy and hepatojejunostomy.

Treatment options for peripheral cholangiocarcinoma

Approximately 10% of CCs arise within the intrahepatic ducts of the liver parenchyma (peripheral or ICC). Unlike hepatocellular carcinoma, peripheral/ICCs are relatively avascular lesions and the hepatic parenchyma does not show cirrhosis. Resection usually requires a lobectomy; the extent depending on the size and location of the tumour(s) and on the presence of satellites. Hepatic resection with negative margins is the only curative treatment option. After complete resection, 5 year survival is 25%, with the majority of recurrences being intrahepatic. Some authors have reported 5 year survival rates of 39–41% but only in patients with negative margins. The clinicopathological factors which predict good overall survival after hepatectomy include absence of clinical symptoms, presence of mucobilia, early-stage tumour, papillary type and postoperative chemotherapy.

Liver transplantation as a primary treatment for ICC remains controversial in view of the limited organ availability and the high recurrence rate. The results of liver transplantation for ICC are however good with 2 and 5 year survival rates of 48% and 23%, respectively, although more than 50% of patients develop a recurrence within 2 years. The best results are obtained in patients with a single lesion and small tumour size. Selected patients with early-stage ICC can be treated with liver transplantation but currently only within clinical trials. Preliminary results with a neoadjuvant chemoradiation protocol have shown a 45% survival rate with a median follow-up >7 years. It is likely that the greatest benefit from neoadjuvant radiochemotherapy and subsequent liver transplantation may be derived by younger patients (<45 years) who develop CC as a result of PSC.

Adjuvant radiotherapy

Adjuvant radiotherapy is used in many centres to sterilize the surgical margins in an attempt to reduce recurrence. Although there are no prospective randomized trials adjuvant radiotherapy (both external beam radiotherapy with or without brachytherapy and intraoperative radiation therapy) is considered standard treatment for patients with CC following resection with curative intent. The results of several retrospective studies seem to confirm benefit of this policy. One study showed a statistically improved 5 year survival rate (39%) in those treated with surgery alone (14%). Others use an adjuvant chemoradiotherapy regimen consisting of postoperative radiotherapy and bolus 5-fluorouracil (5FU) for the first 3 days of each course. Adjuvant chemoradiotherapy is specially indicated in patients with positive resection margins, and/or if regional lymph nodes are involved.

Neoadjuvant treatment

Neoadjuvant chemoradiotherapy is used in selected patients. In this instance, chemoradiotherapy is administered before surgery. Although the reported data are limited, neoadjuvant therapy has been reported to achieve pathological complete response in some patients and to increase significantly the negative margin rates after resection. In patients with stage I or II CC Rea *et al*. reported a 1-, 3- and 5-year survival of 82%, 48% and 21%, respectively, after neoadjuvant radiotherapy. Despite these encouraging reports, the role of neoadjuvant therapy remains unproven and is currently used only in clinical trials.

Palliative treatment for inoperable/incurable disease

This includes photodynamic therapy, stenting, palliative radiotherapy and palliative chemotherapy. The choice depends on local expertise and the condition, motivation and performance status of the patient

Stenting

In patients who are considered inoperable on preoperative assessment and those who are too old and frail or have serious cardiorespiratory disease, palliation of the jaundice is best achieved by percutaneous transhepatic or endoscopic stenting. The endoprosthesis has to be large (8–10 Fr) and may require replacement if it becomes blocked with calcium bilirubinate encrustation. Self-expanding stainless steel wire biliary endoprostheses are now used frequently for these malignant strictures. The big advantage of this type of endoprosthesis over plastic stents is that it can be introduced through a small-calibre sheath. When the latter is removed, the stent expands to achieve an internal diameter much greater than that of conventional plastic endoprostheses. In addition, the wire framework of the stent becomes buried in the wall of the duct and is thus less liable to cause infection and encrustation with calcium bilirubinate. This contributes to their high patency rate, which is reported to be 90%. These self-expanding metal stents are, however, contraindicated in patients with polypoid tumours as the fronds of the neoplasm project through the wire framework into the lumen. In addition, there is a risk of haemorrhage in these cases.

The controversy in the management of inoperable CC has until recently even extended to the endoscopic palliation in respect of the choice of stent (plastic vs metallic). This has now been resolved by a large prospective study of 101 patients who were followed until death or at least 1 year after inclusion into the study. By multivariate analysis, only tumour size predicted survival. Thus, a threshold of 30 mm at diagnosis distinguished two survival profiles: the median survival of patients with a tumour greater than 30 mm was 3.2 months as distinct from 6.6 months for patients with tumour size <30 mm.

The recommended strategy involves the use of metal stents for patients with an inoperable tumour smaller than 30 mm, while larger tumours are efficiently palliated by a plastic stent.

Approximately 15% of patients develop stent occlusion at a mean interval of 4 months after insertion of metal stents. In 80% occlusion is due to tumour overgrowth and in 20% to debris. When the occlusion is caused by tumour ingrowth, the survival is limited to a maximum of 3 months and these patients are best treated by internal plastic stents. Occlusion of metal stents by debris is effectively cleared by sweeping the stent with a balloon catheter.

Photodynamic therapy

Photodynamic therapy with the photosensitizer porfimer and endoscopic direct illumination of the tumour by laser light of 630 nm can provide effective palliation due to recanalization. The photodynamic therapy may be repeated after 6 months. In a randomized study of 39 patients, those receiving photodynamic therapy plus stenting had a longer survival than patients treated by stent insertion alone (16.4 months vs 3.3 months; Ortner 2003). However, photodynamic therapy is only available in some specialized centres.

Palliative radiotherapy

Patients with locally advanced disease without distant metastases are treated with palliative radiotherapy usually with external beam irradiation combined with intraluminal iridium-192 application. Conformational radiotherapy with locoregional chemotherapy is used in some centres. The percutaneous insertion of iridium-192 wire has been used to provide local irradiation (brachytherapy) with good results and reported median survival of 11–23 months. Iridium-192 brachytherapy administered via the transhepatic approach has been shown to provide worthwhile palliation in patients with unresectable disease and after recurrence following previous complete or partial resection. The most common acute complication after brachytherapy is cholangitis followed by duodenal ulceration.

Chemotherapy

Systemic chemotherapy is indicated in patients with a good performance status who are not candidates for locoregional treatment. The evidence on the efficacy of systemic chemotherapy for advanced CC is however limited. 5FU alone or combined with leucovorin results in response rates up to 10%. Combination chemotherapy with the ECF regimen (epirubicin, cisplatin, 5FU) improves the response rate to 40% with a median survival of 11 months. Moderate single-agent activity has been documented for capecitabine and gemcitabine. Both may be administered in combination with leucovorin and infusional 5FU. The combination of gemcitabine with platinum compounds is also effective in patients with good performance status and adequate liver function. In practice, the selection between single agent and combination chemotherapy depends on performance status and motivation of the patient.

Ampullary tumours

Although these are often grouped with periampullary tumours, there is growing evidence based on detailed histochemical and gene expression studies that ampullary tumours form a separate category. They can either be adenomas or adenocarcinomas. The ampulla of Vater encompasses two distinct types of mucosa: intestinal and pancreatobiliary. Hence, ampullary cancers can originate from either and thus constitute two groups: intestinal and pancreatobiliary with distinct immunohistological staining. Published studies suggest that patients with intestinal types tend to have better prognosis than those with pancreatobiliary cancers. Primary ampullary carcinomas are uncommon tumours (six cases per million population) and carry a better overall prognosis than periampullary cancers.

Some ampullary tumours arise in association with familial polyposis coli. Additionally, ampullary carcinomas have a higher resectability rate than periampullary carcinomas, although the prognosis is poor when the disease is advanced. Several studies have now shown that the preoperative staging is best achieved by endoscopic ultrasound (EUS), which has an overall accuracy of tumour (T) staging of 75%, although it is less reliable in detecting lymph node involvement (60%). Its most important contribution, however, is in the detection of pancreatic invasion, for which EUS has an accuracy of 86%, a sensitivity of 83% and a specificity of 87%.

Ampullary cancer is more common in men. Its usual presentation is with painless obstructive jaundice and itching. Physical examination usually reveals a palpable distended gallbladder (Courvoisier's sign) and moderate smooth hepatomegaly.

Treatment

The options for treatment of ampullary tumours include local excision, pancreaticoduodenectomy and endoscopic management (endoscopic papillectomy and debulking, endoscopic stenting). If the lesions can be confidently diagnosed as benign by endoscopic biopsy, treatment is by transduodenal excision of the ampulla of Vater. This can be performed surgically or by endoscopic papillectomy. Local excision is also used for malignant ampullary lesions if:

- the patient is unfit for major surgery
- the tumour is limited to the ampulla of Vater as diagnosed by preoperative EUS (uT1 and UICC-staging pT1); and the histological grading is G1 or G2 without lymphatic infiltration.

Close postoperative follow-up with duodenoscopy and ERCP are necessary in all patients treated by local excision. For all other operable tumours in fit patients the best treatment is obtained by pancreaticoduodenectomy.

Endoscopic management (laser debulking and stenting) is reserved for patients who are unfit for surgery and for advanced inoperable disease.

Periampullary tumours

The periampullary region represents a cross road of three mucosal surfaces: pancreatic, biliary and duodenal mucosa. Excluding the ampulla itself, these tumours (usually carcinomas) can arise from (1) terminal common bile duct, (2) pancreatic duct or (3) adjacent duodenal mucosa. The histochemical distinction between periampullary cancer and ampullary cancer is based on the type of mucin secreted. Thus whereas ampullary cancer tends to secrete sialomucins, periampullary cancer produces sulphated mucins. They are considered to be distinct from primary duodenal carcinoma. By definition these include tumours that are located within 1.0 cm of the papilla. Periampullary cancers are generally more advanced than ampullary cancer at presentation. Their symptoms include jaundice and pruritus, anorexia, nausea, vomiting or weight loss. Many patients also complain of abdominal pain. Some also complain of diarrhoea (steatorrhoea) due to malabsorption of fat consequent on the absence of lipase.

The staging of periampullary cancer is based on the TNM system giving the following stage grouping:

- stage 0: TisN0M0
- stage I: T1N0M0
- stage II: T2–3N0M0
- stage III: T1–3N1M0
- stage IV: T1–4N0-1M1

Treatment

When operable and in fit patients, the treatment is excision by pancreaticoduodenectomy. Local excision (transduodenal papilloduodenectomy) is indicated in frail patients. The resection includes the papilla, the distal bile duct and the pancreatic duct with surrounding pancreatic tissue. Following excision, the pancreatic and common bile duct are sutured together and then to the defect in the medial wall of the duodenum. The reported experience with this procedure is limited but the operation appears safe.

Results after radical resection of periampullary cancer (pancreaticoduodenectomy) have been improving with 5 year survival rates in recent years ranging from 20% to 61% with a median of 35%. Furthermore, the reported postoperative mortality rate has fallen to single figures in high-volume centres. However, only 63% of patients have curative (R0) resection and 50% have involved lymph nodes, most commonly the posterior pancreaticoduodenal nodal group. Survival after surgical resection is related to the extent of local invasion of the primary lesion, lymph node involvement, vascular invasion, perineural invasion, cellular differentiation, and uninvolved surgical margins.

Parasitic infestations of the biliary tract

In addition to infestations with schistosomes and the larvae of *Taenia echinococcus*, the liver and biliary tract is involved with various other parasitic disorders. Some infestations, such as toxocariasis, remain subclinical in the majority of cases. Children usually acquire infection with *Toxocara canis* from their pets. In addition to hepatic granulomas, there may be central nervous system and eye involvement. The latter is serious and leads to endophthalmitis which can be mistaken for retinoblastoma in these children.

The parasitic disorders which are of significance in surgery of the biliary tract include infestation with *A. lumbricoides, C. sinensis* and *Fasciola hepatica*.

Ascaris lumbricoides

Infestation with this nematode is endemic and prevalent in Asia, China and Africa. It is also found in the rural areas of Europe, the USA and Latin America. The adult worms live in the upper reaches of the small intestine but migrate to and from the bile duct through the ampulla of Vater, up the oesophagus and down to the appendix. The ova are excreted in the stools of infected individuals and contaminate soil and vegetables. Following ingestion of the encysted larvae and dissolution of the cyst wall by the gastric juice, the free larvae penetrate the intestinal mucosa to reach the portal venous system and thus the liver or the lung via the intestinal lymphatics and the thoracic duct. The pulmonary larvae are carried from the alveoli to the pharynx and then swallowed to reach the upper part of the small intestine where they mature into adult worms. The larval migration may involve other organs, e.g. central nervous system and kidneys.

The majority of adult worms migrating into the biliary tract die after a few weeks and may form a nidus for stone formation. Secondary infection of the bile with *E. coli* and other enteric organisms is common and is thought to play a role in the formation of calcium bilirubinate stones and the development of CC. Pyogenic liver abscess may complicate the disease. The usual hepatic lesions are granulomas surrounding the ova, which are deposited in the smaller bile ducts.

Clinical features

The stage of larval migration is accompanied by systemic symptoms: rigors, generalized aches, malaise, cough and asthmatic attacks. Eosinophilia is invariably present. Migration of the adult worms into the bile duct induces episodes of pain in the epigastric region. Jaundice is encountered in 20%. *A. lumbricoides* infestation is one of the commonest causes of jaundice in children and young adults in Africa. At this stage, the clinical picture is usually dominated by recurrent attacks of cholangitis due to calculus formation and secondary infection.

Treatment

Surgical intervention is necessary to deal with the biliary complications. An initial laparotomy is required to look for and remove (by enterotomy) adult worms which can be easily palpated through the intact bowel wall. Exploration of the bile duct, and removal of stones and worms, is performed next. A completion cholangiogram or, preferably, an inspection with the choledochoscope is advisable at the end of the duct exploration. T-tube drainage is essential in all cases and should be carried out with a large tube (16–18 Fr) to enable subsequent percutaneous stone and parasite extraction through the T-tube tract, if necessary. Surgical intervention should be followed by antihelminthic therapy.

Clonorchis sinensis

Man is the definitive host of this trematode which is widely distributed in China and East Asia. The adult worms live in the biliary tract and occasionally in the pancreatic duct. The eggs are excreted in the faeces and ingested by freshwater snails (e.g. *Parafossarulus manchouricus*) where they develop into cercariae. The free-swimming cercariae then penetrate freshwater fish and encyst themselves in the muscles of the host as metacercariae. Man becomes infected by eating contaminated fish, the encysted metacercariae being released in the duodenum. They then migrate into the biliary tract via the ampulla of Vater and mature into adult worms. These cause dilatation of the biliary tract with fibrosis of the ducts and adenomatous bile duct hyperplasia. Secondary infection of the biliary tract with enteric organisms is extremely common and results in death of the worm and stone formation. Recurrent episodes of cholangitis and septicaemia, and the development of CC, account for the appreciable mortality of this parasitic infestation. The diagnosis is confirmed by the demonstration of the typical ova in the faeces.

Treatment

Mild cases can be managed conservatively with chloroquine (300 mg for 2–6 months). Surgical intervention is needed for jaundice and cholangitis. In addition to removal of worms and calculi from the bile duct, an internal biliary drainage (choledochoduodenostomy/jejunostomy or sphincteroplasty) is performed.

Fasciola hepatica

This is primarily an infestation of sheep and cattle. Man is an accidental host, acquiring the disease by eating wild watercress contaminated with metacercariae. The disease has a worldwide distribution and is common in Latin America and the USA. In the UK outbreaks of the disease have occurred in Hampshire and the Lake District at Silverside. The ova are excreted in the faeces of infected animals, develop into miracidia in freshwater and subsequently colonize the intermediate host, a freshwater snail (*Lymnaea truncatula*). Within this host, they mature through various stages into metacercariae which become encysted on neighbouring water plants. Subsequent to ingestion, the free metacercariae are released in the upper part of the small intestine and penetrate the bowel to reach the peritoneal cavity. They then migrate across the peritoneal cavity and enter the liver parenchyma after penetration of the liver capsule. Maturation occurs in the bile ducts. Migration of the metacercariae may occur to other organs: kidneys, muscles, brain and subcutaneous tissue.

Clinical features

The disease is often asymptomatic if the infestation is mild. Systemic symptoms signify heavy infestations and include malaise, anorexia, nausea, vomiting, fever and weight loss. An urticarial rash, jaundice or hepatosplenomegaly may develop. Diagnosis is confirmed by the demonstration of ova in the stool.

Treatment

The disease is usually treated medically with bithionol (50 mg daily for 3 weeks). Surgical intervention is indicated only in cases of biliary obstruction and cholangitis.

GUIDE TO FURTHER READING

Adamsen, S, Hansen OH, Funch Jensen P, *et al.* Bile duct injury during laparoscopic cholecystectomy: a prospective nationwide series. *J Am Coll Surg* 1997;**184**:571–578.

Alberts SR, Al-Kbatib H, Mahoney MR, *et al.* Gemcitabine, 5-fluorouracil and leucovorin in advanced biliary and gallbladder cancer: A North Central Cancer Treatment Group Phase II trial. *Cancer* 2005;**11**:111–118.

Allemann P, Schafer M, Demartines N. Critical appraisal of single port access cholecystectomy. *Br J Surg* 2010; **97**:1476–1480.

Anne M, Covey AM, Brody LA, *et al.* Incidence, patterns, and clinical relevance of variant portal vein anatomy. *AJR Am J Roentgenol* 2004;**183**:1055–1063.

Arora DS, Ramsdale J, Lodge JP, Wyatt JI. p53 but not bcl-2 is expressed by most cholangiocarcinomas: a study of 28 cases. *Histopathology* 1999;**34**:497–501.

Berdah SV, Delpero JR, Garcia S, *et al.* A western surgical experience of peripheral cholangiocarcinoma. *Br J Surg* 1999;**83**:1517–1521.

Bergquist A, Glaumann H, Persson B, *et al.* Risk factors and clinical presentation of hepatobiliary carcinoma in patients with primary sclerosing cholangitis: a case-control study. *Hepatology* 1998;**27**:311–316.

Berr F. Photodynamic therapy for cholangiocarcinoma. *Semin Liver Dis* 2004;**24**:177–187.

Bismuth H. Postoperative strictures of the bile ducts. In: Blumgart LH (ed.) *The Biliary Tract V.* New York, NY: Churchill-Livingstone, 1982:209–218.

Born P, Rösch T, Brühl K, *et al.* Long-term results of endoscopic and percutaneous transhepatic treatment of benign biliary strictures. *Endoscopy* 1999;**31**:725–731.

Bret PM, Reinhold C. Magnetic resonance cholangiopancreatography. *Endoscopy* 1997;**29**:472–486.

Carey MC, Cahalane MJ. Whither biliary sludge? *Gastroenterology* 1988;**95**:508–523.

Carrol BJ, Birth M, Phillips EH. Common bile duct injuries during laparoscopic cholecystectomy that result in litigation. *Surg Endosc* 1998;**12**:310–314.

Cetta E. The role of bacteria in pigment gallstone disease. *Ann Surg* 1991;**213**:315–326.

Cetta E. Do surgical and endoscopic sphincterotomy prevent or facilitate recurrent common duct stone formation? *Arch Surg* 1993;**128**:329.

Chartrand-Lefebvre C, Dufresne MP, Lafortune M, *et al.* Iatrogenic injury to the bile duct: a working classification for radiologists. *Radiology* 1994;**193**:523–526.

Chijiiwa K, Tanaka M. Indications for and limitations of extended cholecystectomy in the treatment of carcinoma of the gall bladder. *Eur J Surg* 1996;**162**:211–216.

Coakely FV, Schwartz LH. Magnetic resonance cholangiopancreatography. *J Magn Reson Imaging* 1999;**9**:157–162.

Cuschieri A, Croce E, Faggioni A, *et al.* EAES multicentre prospective randomized trial comparing two-stage vs single-stage management of patients with gallstone disease and ductal calculi. *Surg Endosc* 1999;**13**:952–957.

Davidoff AM, Pappas TN, Murray EA, *et al.* Mechanisms of major biliary injury during laparoscopic cholecystectomy. *Ann Surg* 1992;**215**:196–202.

Deziel DJ. Complications of cholecystectomy: incidence, clinical significance and diagnosis. *Surg Clin North Am* 1994;**74**:809–823.

Eltawil K, Gustafsson B, Kidd M, Modlin I. Neuroendocrine tumors of the gallbladder: an evaluation and reassessment of management strategy. *J Clin Gastroenterol* 2010;**44**:687–695.

Farrell RJ, Noonan N, Khan I, *et al.* Carcinoma of the ampulla of Vater: a tumour with a poor prognosis? *Eur J Gastroenterol Hepatol* 1996;**8**:139–144.

Fava G, Marzioni M, Glaser S, *et al.* Molecular pathology of biliary tract cancers. *Cancer Lett* 2007;**250**:155–167.

Flanebaum L, Aldsen SM, Trooskin SZ. Use of cholescintigraphy with morphine in critically ill patients with suspected cholecystitis. *Surgery* 1989;**106:**668–674.

Fletcher DR, Hobbs MS, Tan P, *et al.* Complications of cholecystectomy: risks of the laparoscopic approach and protective effects of operative cholangiography: a population-based study. *Ann Surg* 1999;**229**:449–457.

Flowers JL, Zucker KA, Graham SM, *et al.* Laparoscopic cholangiography: results and indications. *Ann Surg* 1992;**215**:209–216.

Greenen DJ, Greenen JE, Hoagen WJ, *et al.* Endoscopic therapy for benign bile duct strictures. *Gastrointest Endosc* 1989;**35**:367–371.

Gupta N, Solomon H, Fairchild R, Kaminski DL. Management and outcome of patients with combined bile duct and hepatic artery injuries. *Arch Surg* 1998;**133**:176–181.

Hemming AW, Reed AI, Fujita S, *et al.* Surgical management of hilar cholangiocarcinoma. *Ann Surg* 2005;**241**:693–699.

Jonas S, Neuhaus P. The perspective of liver transplantation for cholangiocarcinoma. *Liver Transpl* 2007;**13**:1356–1361.

Kadesky KM, Schopf B, Magee JF, Blair GK. Proximity injury by ultrasonically activated scalpel during dissection. *J Paediatr Surg* 1997;**32**:878–879.

Kern KA. Medicolegal perspective of laparoscopic bile duct injuries. *Surg Clin North Am* 1994;**74**:979–984.

Kiviluoto T, Siren J, Luukkonen P, Kivilaakso E. Randomised trial of laparoscopic versus open cholecystectomy for acute and gangrenous cholecystitis. *Lancet* 1998;**351**:321–325.

Klein P, Reingruber B, Kastl S, *et al.* Is local excision of pTI-ampullary carcinomas justified? *Eur J Surg Oncol* 1966;**22**:366–371.

Kubo H, Chijiiwa Y, Akahoshi K, *et al.* Pre-operative staging of ampullary tumours by endoscopic ultrasound. *Br J Radiol* 1999;**72**:443–447.

Kullman E, Borch K, Lindstrom E, *et al.* Value of routine intraoperative cholangiography in detecting aberrant bile ducts and bile duct injuries during laparoscopic cholecystectomy. *Br J Surg* 1996;**83**:171–175.

Lai PB, Kwong KH, Leung KL, *et al.* Randomized trial of early versus delayed laparoscopic cholecystectomy for acute cholecystitis. *Br J Surg* 1988;**85**:764–767.

Leung J, Guiney M, Das R. Intraluminal brachytherapy in bile duct carcinomas. *Aust NZ J Surg* 1996;**66**:74–77.

Lo CM, Fan ST, Liu CL, *et al.* Early decision for conversion of laparoscopic to open cholecystectomy for treatment of acute cholecystitis. *Am J Surg* 1997;**173**:513–517.

Lo CM, Liu CL, Fan ST, *et al.* Prospective randomized study of early versus delayed laparoscopic cholecystitis. *Ann Surg* 1998;**227**:461–467.

Lorimer JW, Fairfull Smith RJ. Intraoperative cholangiography is not essential to avoid duct injuries during laparoscopic cholecystectomy. *Am J Surg* 1994;**169**:344–347.

MacFadyen BV, Vecchio R, Ricardo AE, Mathis CR. Bile duct injury after laparoscopic cholecystectomy: the United States experience. *Surg Endosc* 1998;**12**:315–321.

Marks J, Tacchino R, Roberts K, *et al.* Prospective randomized controlled trial of traditional laparoscopic cholecystectomy versus single-incision laparoscopic cholecystectomy: report of preliminary data. *Am J Surg* 2011; **201**:369–372.

Martin RF, Rossi RL. Bile duct injuries: spectrum, mechanism of injury, and their prevention. *Surg Clin North Am* 1994;**74**:781–803.

Meyers WC, and Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. *N Engl J Med* 1994;**324**:1073–1078.

Miyazaki M, Itoh H, Ambiru S, *et al.* Radical surgery for advanced gallbladder carcinoma. *Br J Surg* 1996;**83**:478–481.

Morita M, Ueno E, Masukawa A, *et al.* Defining juxtapapillary diverticulum with 3D segmented trueFISP MRCP: comparison with conventional MRCP sequences with an oral negative contrast agent. *Jpn J Radiol* 2009;**27**:423–429.

Nichols DM, Macleod AJ. Choledocholithiasis associated with malignant biliary obstruction – significance and management. *Clin Radiol* 1998;**53**:49–52.

Ogura Y, Kawarada Y. Surgical strategies for carcinoma of the hepatic duct confluence. *Br J Surg* 1998;**85**:20–24.

Parc Y, Frileux P, Balladur P, *et al.* Surgical strategy for the management of hilar bile duct cancer. *Br J Surg* 1997;**84**:1675–1679.

Prat F, Chapat O, Ducot B, *et al.* Predictive factors for survival of patients with inoperable malignant distal biliary strictures: a practical management guideline. *Gut* 1998;**42**:76–80.

Richardson MC, Bell G, Fullarton GM and the West of Scotland Laparoscopic Cholecystectomy Audit Group. Incidence and nature of bile duct injuries following laparoscopic cholecystectomy: an audit of 5913 cases. *Br J Surg* 1996;**83**:1356–1360.

Roa JC, Tapia C, Cakir A, *et al.* Squamous cell and adenosquamous carcinomas of the gallbladder: clinicopathological analysis of 34 cases identified in 606 carcinomas. *Modern Pathol* 2009;**24**:1069–1078.

Robertson AJ, Rela M, Karani J, *et al.* Laparoscopic cholecystectomy injury: an unusual indication for liver transplantation. *Transpl Int* 1998;**11**:449–451.

Russel JC, Walsh SJ, Mattie AS, Lynch JT. Bile duct injuries, 1989–93: a statewide experience. *Arch Surg* 1996;**131**:382–388.

Sasaki A, Aramaki M, Kawano K, *et al.* Intrahepatic peripheral cholangiocarcinoma: mode of spread and choice of surgical treatment. *Br J Surg* 1998;**85**:1206–1209.

Savader SJ, Lillemoe KD, Prescott CA, *et al.* Laparoscopic cholecystectomy-related bile duct injuries: a health and financial disaster. *Ann Surg* 1997;**225**:268–273.

Schol FP, Go PM, Gouma DJ. Outcome of 49 repairs of bile duct injuries after laparoscopic cholecystectomy. *World J Surg* 1995;**19**:753–756.

Sithithaworn P, Haswell-Elkins MR, Mairianq P, *et al.* Parasite- associated morbidity: liver fluke infection and bile duct cancer in northeast Thailand. *Int J Parasitol* 1994;**24**:833–843.

Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies performed by Southern US surgeons. *N Engl J Med* 1991;**324**:1073–1078.

Soyer P, Brouland JP, Boudiaf M, *et al.* Color velocity imaging and power Doppler sonography of the gall bladder wall: a new look at sonographic diagnosis of acute cholecystitis. *AJR Am J Roentgenol* 1998;**171**:183–188.

Stewart L, Way LW. Bile duct injuries during laparoscopic cholecystectomy. Factors that influence the results of treatment. *Arch Surg* 1995;**130**:1123–1128.

Strasberg SM, Herd M, Soper NJ. An analysis of the problem of biliary injury during laparoscopic cholecystectomy. *J Am Coll Surg* 1995;**180**:101–125.

Tagle FM, Lavergne J, Barkin JS, Unger SW. Laparoscopic cholecystectomy in the elderly. *Surg Endosc* 1997;**11**:636–638.

Tamada K, Yasuda Y, Nagai H, *et al.* Limitation of cholangiography in assessing longitudinal spread of extrahepatic bile duct carcinoma to the hepatic side. *J Gastroenterol Hepatol* 1999;**14**:691–698.

Taragona EM, Perez Ayuso RM, Bordas JM, *et al.* Randomised trial of endoscopic sphincterotomy with gallbladder left *in situ versus* open surgery for common bile duct calculi in high-risk patients. *Lancet* 1996;**347**:926–929.

Tokunaga Y, Nakayama, N, Ishikawa Y, *et al.* Surgical risks of acute cholecystitis in the elderly. *Hepatogastroenterology* 1997;**44**: 671–676.

Uggowitzer M, Kugler C, Schramyer G, *et al.* Sonography of acute cholecystitis: comparison of color and power Doppler sonography in detecting a hypervascularized gallbladder wall. *Am J Roentgenol* 1997;**168**:707–712.

van Gulik TM, Rauws EA, Gonzalez Gonzalez D, *et al.* Pre- and post-operative irradiation in the treatment of resectable Klatskin tumours. *Ned Tijdschr Geneeskd* 1997;**141**:1331–1337.

Wang YJ, Lee SD, Shyu JK, Lo KJ. Clinical experience in 126 patients with tissue-proved proximal cholangiocarcinoma. *J Gastroenterol Hepatol* 1994;**9**:134–137.

Watanapa P. Cholangiocarcinoma in patients with opisthorchiasis. *Br J Surg* 1996;**83**:1062–1064.

Wiesner W, Beglinger Ch, Oertli D, Steinbrich W. Juxtapapillary duodenal diverticula: MDCT findings in 1010 patients and proposal for a new classification. *JPR-BTR* 2009;**92**:191–194.

Wilson T, Polly L, Wai-ki L, *et al.* Primary hepatolitiasis. *Adv Anat Pathol* 2011;**18**:318–328.

CHAPTER 26

Disorders of the spleen and lymph nodes

SIR ALFRED CUSCHIERI

Introduction

The spleen and lymph nodes combined constitute the bulk of the reticuloendothelial (monocyte–macrophage) system. During fetal development, the spleen serves a transient role of haematopoiesis. In adult life the spleen and lymph nodes share many functions with regard to immune processing, and are, therefore, not surprisingly, commonly involved in similar disease processes.

Surgical anatomy of the spleen

Embryology

The spleen appears at the fifth week of intrauterine life as mesodermal thickenings in the coelomic epithelium of the left leaf of the dorsal mesogastrium. These proliferating areas undergo condensation and vascularization forming distinct tissue aggregates which then fuse to form the organ but with retention of the segmental nature of the adult spleen.

The development of either a ‘compact’ or ‘diffuse’ multinotched organ is determined by the initial number of the mesenchymal masses. Reticular cells provide the connective tissue support of the fetal spleen and contribute to the white splenic pulp. The formation of the white pulp is associated with an influx of lymphoid cells. These are primarily B lymphocytes and express surface immunoglobulin (Ig)G and IgD. Follicular dendritic cells are also found with the –cells. The initial –cell accumulation develops in the marginal zone, separating red from white pulp. The reticular cells form sheaths around the central arterioles of the red pulp and ultimately form the periarteriolar lymphatic sheath (PALS). At 19 weeks further lymphoid cells accumulate around the central arteriole. These represent precursor T-cells and interdigitating reticular cells. At 23 weeks, –cells and follicular dendritic reticulum cell precursors are present in the primary follicles of PALS. The spleen increases its size during intrauterine life until birth, when it starts to shrink from reduction of the erythrocyte reservoir of the red pulp.

Anatomy of the adult spleen

The spleen is normally situated in the left upper quadrant tucked under and against the left dome of the diaphragm and overlaid by the lower ninth to eleventh left ribs such that it is impalpable and has to enlarge two to three times before it becomes palpable on examination. It is overlapped anteriorly by the fundus of the stomach with its long axis lying along the line of the tenth rib. The adult spleen has an anterior and posterior extremity (referred to as superior and inferior poles respectively), superior and inferior border and a visceral and diaphragmatic surface. The visceral surface is intimately related to four abdominal organs: stomach (gastric impression), tail of the pancreas, left kidney (renal impression) and the splenic flexure (colic impression).

The gastric impression is directly related to the posterior wall of the stomach from which it is separated by the lesser sac. The spleen is intimately related to the tail of the pancreas as this lies in the lienorenal ligament and is liable to injury unless care is taken during splenectomy (Figure 26.1). Although usually located in the left hypochondrium, abnormal locations are well documented, including the retrorenal space, pelvis and even the scrotum.

The normal dimensions of the spleen are 13 cm in length (distance between the superior and inferior pole), with a range 9–16 cm, breadth (distance from the anterior to posterior border) of 9 cm with a range of 6–11 cm, and width (distance between the diaphragmatic to the visceral surface at the widest point) of 3 cm. Hence size varies considerably between healthy adult individuals. Three distinct morphological types of the human adult spleen are recognized: crescentic, triangular and rhomboid; however, the shape can change with enlargement of the organ by disease. Crescentic spleen refers to organs with regular hila extending from prominent superior to inferior poles. Triangular spleens consist of organs with three borders (anterior, posterosuperior and posteroinferior) and three poles (superior, inferior and posterior). In rhomboid spleens, the poles are expanded into borders, particularly the inferior pole, producing an organ with four unequal sides.

The spleen is attached to the diaphragm and retroperitoneal organs by a fascia and peritoneal folds: splenocolic, diaphragmatic and lienorenal. The attachment to the fundus of the stomach is via the gastrosplenic section of the greater omentum that contains the short gastric vessels. Especially in obese subjects, the spleen is often covered with greater omentum, which indeed may be adherent to the surface of the organ. The splenic substance consists of red pulp (80%) made up of sinuses and sinusoids and cellular cords containing macrophages and white pulp consisting of lymphoid tissue (20%). The red pulp is concerned with maturation and the removal of damaged and senescent red cells, whereas the white pulp forms a major component of the cellular immune surveillance and protection of the body.

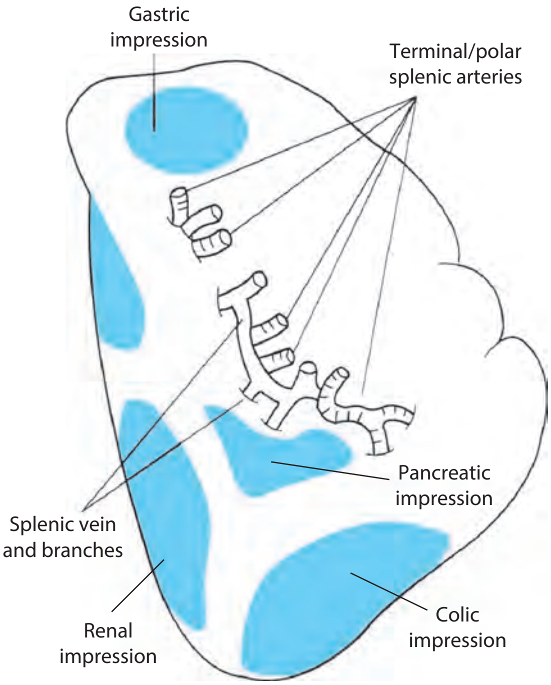


Figure 26.1 Visceral surface and hilum of the spleen.

Splenic vasculature and segmentation

The main arterial supply is from the common splenic artery (origin from the coeliac axis) which is tortuous and runs along the upper border of the pancreas and is anatomically divisible into four segments: suprapancreatic (longest segment), pancreatic (looped or coiled segment lying in a groove on the superior border of pancreas), prepancreatic (commonest site of division of the splenic artery particularly with the distributed vascular pattern) and prehilar segment (lies between the tail of the pancreas and splenic hilum).

The surgically significant morphological distinction in the blood supply to the spleen is between 'compact' (magisteral) and 'diffuse (distributed)' as this determines the nature of the terminal branching of the splenic artery (Figure 26.2a,b).

The division of the common splenic artery into its two terminal arteries (superior and inferior) occurs at a variable distance from the spleen. The two terminal arteries give off the segmental arteries to the spleen. The central segmental arteries always arise from the two terminal arteries, but the segmental arteries to the poles of the spleen (polar) have very variable origin (directly from the common splenic artery, from the terminal arteries and, in the case of the lower polar artery, from the gastroepiploic artery). The segmental vasculature nature of the spleen is important in relation to partial splenectomy and to splenic preservation. The splenic segments, which are separated by avascular planes, are usually four in number: two central and two polar.

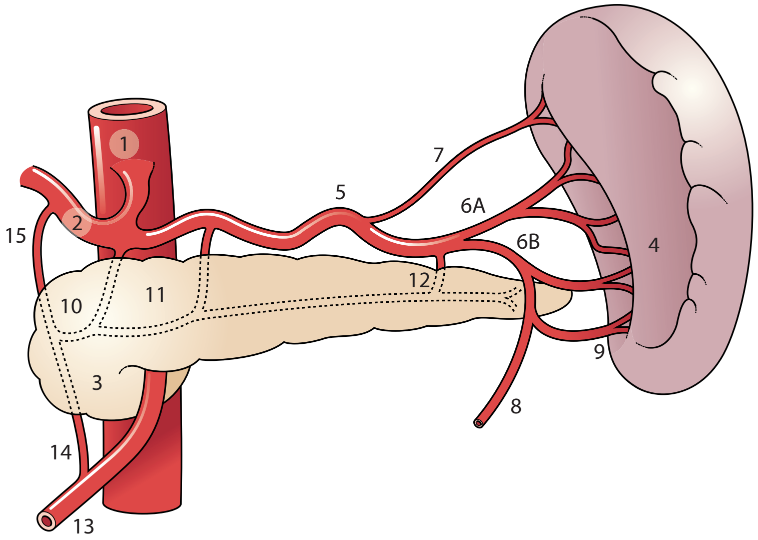
The venous drainage mirrors the arterial supply with the splenic vein joining the superior mesenteric vein behind the neck of the pancreas to form the portal vein. The spleen also drains through the short gastric veins and these assume importance in two situations. In the first instance, they become varicose and lead to gastric varices when there is splenic vein thrombosis. Second, they provide the basis of transsplenic decompression of oesophageal varices by the distal splenorenal shunt (DSRS) of Warren. The microcirculation of the spleen is unique and has two systems: closed and open circulation (Figure 26.3).

Segmentation

The splenic segments are anatomically distinct areas of splenic tissue, identified by corrosion casting and numbering 3–7 with a mean of 4, which extend from the anterior to posterior borders perpendicular to the long axis of the spleen. Each segment has its own arterial and venous territory, but two morphological types are recognized: central and polar, with the central segments being wedge-shaped and larger than the polar segments. The polar segments are pyramidal and located at either end of the spleen. The individual splenic segments are separated by a relatively avascular plane and are important in conservative splenic surgery.

Accessory spleens

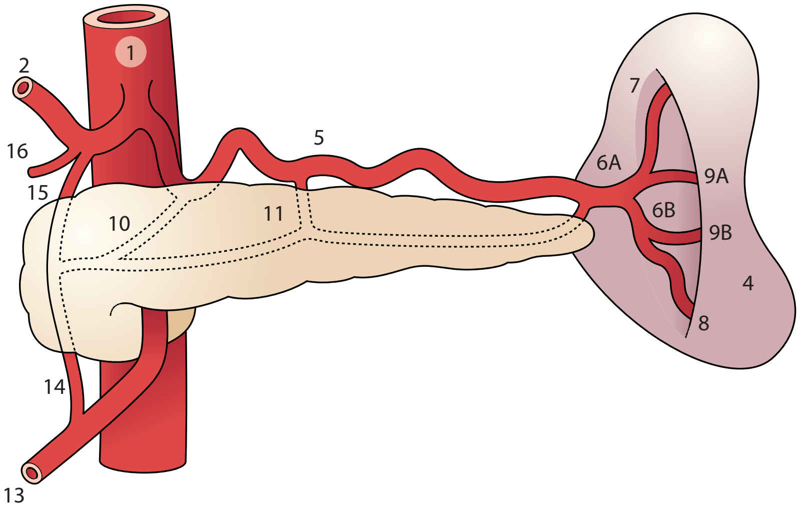
These are also known as splenunculi and are present in 10% of adults (Figure 26.4). They can be multiple but rarely exceed 10 and



(a)

1 Aorta
2 Hepatic artery
3 Pancreas
4 Spleen (diffuse, multi-notched)
5 Common splenic artery
6 A Superior and
  B Inferior splenic arteries
7 Superior polar artery

8 Left gastro-epiploic artery
9 Inferior polar artery
10 Dorsal pancreatic artery
11 Arteria pancreatica magna
12 Arteria caudis pancreatica
13 Superior mesenteric artery
14 Inferior pancreatic-duodenal artery
15 Superior pancreatic-duodenal artery



(b)

1 Aorta
2 Hepatic artery
3 Pancreas
4 Spleen (compact)
5 Common splenic artery
6 A Superior and
  B Inferior splenic arteries
7 Superior polar artery

8 Inferior polar artery
9 A/B Central segmental vessels
10 Dorsal pancreatic artery
11 Arteria pancreatica magna
12 Arteria caudis pancreatica
13 Superior mesenteric artery
14 Inferior pancreatic-duodenal artery
15 Superior pancreatic-duodenal artery
16 Gastroduodenal artery

Figure 26.2 (a) Splenic artery blood supply to the pancreas and spleen (distributed type). (b) Splenic artery blood supply to the pancreas and spleen (magisteral type).

most are situated near the hilum (tail of the pancreas, gastrosplenic ligament, lienorenal ligament and greater omentum). However, they can be located in other sites including the omentum along the greater curvature of the stomach, the mesenteries of the small and large intestine, the broad ligament of the uterus and the pouch of Douglas. They have even been reported in the left testis. The various sites for accessory spleens are grouped as follows:

- perisplenic area: hilum, splenic vascular pedicle, tail of pancreas
- greater omentum: along the greater curvature of the stomach
- mesenteries: small and large bowel
- pelvis/groin: left broad ligament, pouch of Douglas, left testis.

There is some evidence that accessory spleens decline in frequency with advancing years presumably secondary to atrophy, and are thus rare in elderly people. Accessory spleens probably result from failure or incorporation of subsegments during embryological development. They have an organized splenic architecture and separate arterial supply, usually from the inferior polar artery. Accessory spleens are important during splenectomy for haematological conditions such as immune thrombocytopenic purpura, for if left behind they may hypertrophy and cause recurrence of the disease. There is some evidence that accessory spleens are less easily identified during laparoscopic splenectomy.

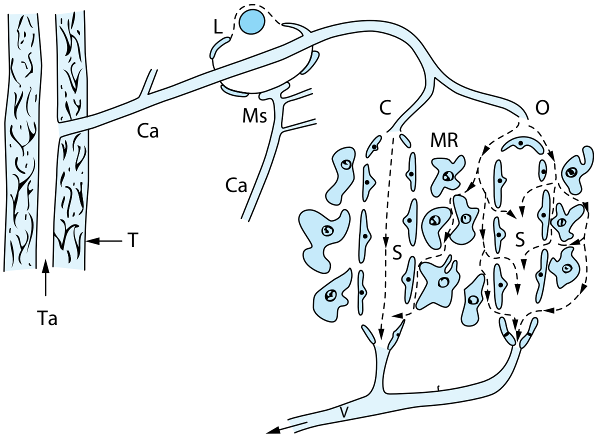


Figure 26.3 Splenic microcirculation: T, trabeculum; Ta, trabecular artery; Ca, central artery; L, lymphoid tissue; Ms, marginal sinus; C, closed circulation; O, open circulation; MR, macrophages and reticulum of red pulp; S, sinus; v, collecting vein.

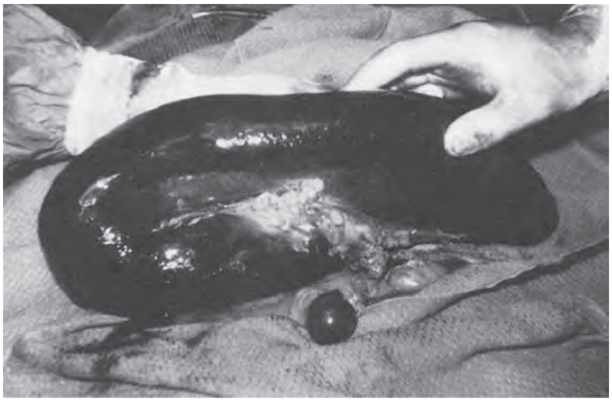


Figure 26.4 Accessory spleen during open splenectomy.

Histology, cytology and blood flow through the spleen

The spleen is covered by a capsule which consists of two layers, an external serosal and an internal fibroelastic coat, which extend as trabeculae of mainly fibrous tissue containing some muscle cells. The trabeculae blend with a reticular internal framework forming the stroma of the organ. The splenic pulp is composed of two distinct tissues: the *red pulp*, which occupies 80% and is concerned with elimination of effete red cells, and the white pulp, containing lymphoid tissue and accounting for the remaining 20%. A schematic diagram of the of the red and white pulp of the spleen is shown in Figure 26.5.

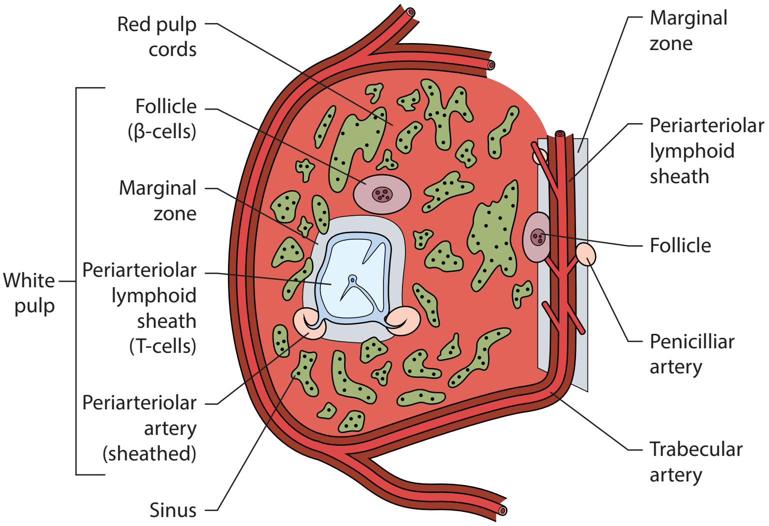


Figure 26.5 Histological appearance of the red and white pulp of the spleen. (Adapted from Buckley.)

Table 26.1 Cellular components of the spleen

Tissue	Cell type
Red pulp	
Cords	Macrophages Reticular cells
Sinuses	Lymphocytes Interendothelial cells
White pulp	
Periarteriolar lymphoid sheaths	T lymphocytes (central) B lymphocytes (peripheral)
Follicles	Dendritic cells Tingible body Macrophages B lymphocytes
Marginal zone	B lymphocytes Marginal zone macrophages

The spleen together with lymph nodes, tonsils, Peyer's patches, etc., constitute the secondary lymphoid organs of the body as distinct from the primary lymphoid organs (thymus and bone marrow) which are the site of lymphopoiesis. Lymphocytes and antigen-presenting cells are found in all the secondary lymphoid organs. The cellular components of the spleen are shown in Table 26.1.

Functions of the spleen and lymph nodes

Haematological functions of the spleen

Because of the peculiar anatomical arrangements of its blood vessels, the spleen is ideally suited as a site of 'quality control' of the erythrocyte population. It removes fragmented, damaged or senescent red cells from the circulation – a process known as culling. It also plays a role in the remodelling of the surface of maturing erythrocytes and in preserving the normal relationship between their membrane surface area and volume. Target cells, which have a relatively high ratio of membrane to intracellular haemoglobin, appear in the peripheral blood soon after splenectomy. A variety of intraerythrocyte inclusions are removed by the spleen by a process known as pitting, after which the red cells are returned to the circulation. Among the inclusions removed are Howell–Jolly bodies, which are probably nuclear remnants (Figure 26.6), siderotic granules (haemosiderin aggregates laid down during normal erythroid maturation) and Heinz bodies, which are pathological aggregates of denatured haemoglobin. Thus after splenectomy, Howell–Jolly bodies and siderotic granules may be seen in the peripheral blood, and red cells show striking changes in shape and size, including the appearance of acanthocytes, irregularly crenated cells and target forms.

The spleen is very effective in the clearance of particulate matter from the circulation – an important function for the timely immune response to bloodborne antigens. The human spleen, contrary to that of other animals, holds relatively little blood in relation to the circulating blood volume, and as such has no significant role in blood storage. Within its volume, however, is a large number of sequestered platelets. Following splenectomy, there is a transient thrombocytosis that can lead to a clinically significant hypercoagulable state.

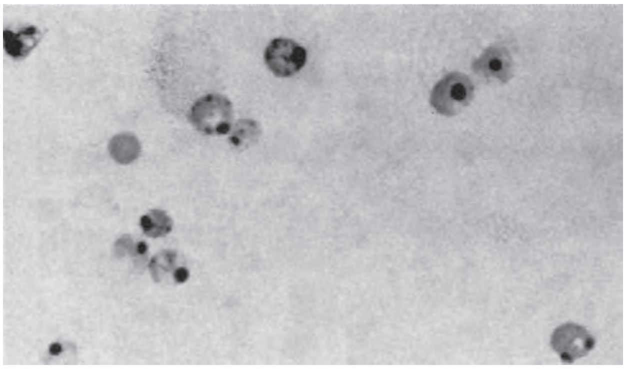


Figure 26.6 Blood film after removal of the spleen. The red cells contain ribonucleoprotein granules (Howell–Jolly bodies) which are normally pitted out by the spleen (Giemsa stain × 680).

The spleen is involved with haemopoiesis only in fetal life, with virtually no blood formation in the organ after birth. However, there may be a reversion to this fetal pattern of erythropoiesis in certain disease states and it is thought that the spleen may become an important organ of red cell production in at least some patients with progressive fibrosis of the bone marrow, i.e. myelosclerosis (Figure 26.7). A variable amount of splenic haematopoiesis also occurs in children with congenital haemolytic anaemia.

Immunological functions of the spleen and lymph nodes

The basic arrangement of lymphoid tissue in spleen and lymph nodes is shown diagrammatically in Figure 26.8. Each population of lymphocytes is in constant flux, with a continuous recirculation of lymph into the bloodstream at the thoracic duct. About one-half of the small lymphocytes of the spleen are migratory. Approximately one-quarter of the body's entire population of T-cells reside within the spleen at any point in time.

Lymphocytes enter the lymph nodes through the permeable walls of the postcapillary (epithelioid) venules of the paracortex; in the spleen, the site of transit is the marginal sinuses bordering the Malpighian corpuscles. T-cells tend to congregate in the paracortex of lymph nodes and form a periarteriolar lymphoid sheath in the spleen. –cells congregate between the T-cells to form small clusters, or primary follicles, at the periphery of the outer cortex of nodes or in the spleen, adjacent to the marginal sinuses. A humoral response following antigenic stimulation involves co-operation between T- and –cells possibly at the site of antigen localization on the surface of large dentritic cells. Immunoglobulin-synthesizing cells appear within days in the medullary cords of lymph nodes and in the red pulp of the spleen. Germinal centres, or secondary follicles, later appear within the primary follicles, and reach their maximum development about 8 weeks following antigenic stimulation. Mature lymphocyte populations exhibit numerous mitotic figures, and are enlarged and plump in comparison with senescent ones.

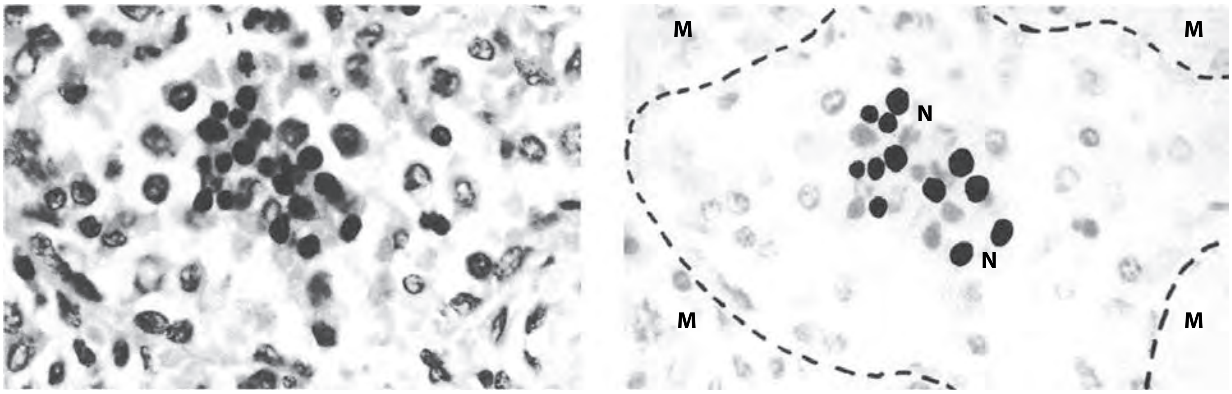


Figure 26.7 An erythropoietic island within a sinusoid in the red pulp of the spleen from a patient with myelosclerosis. N, normoblast; ---, wall of sinus; M, red pulp meshwork surrounding the sinus.

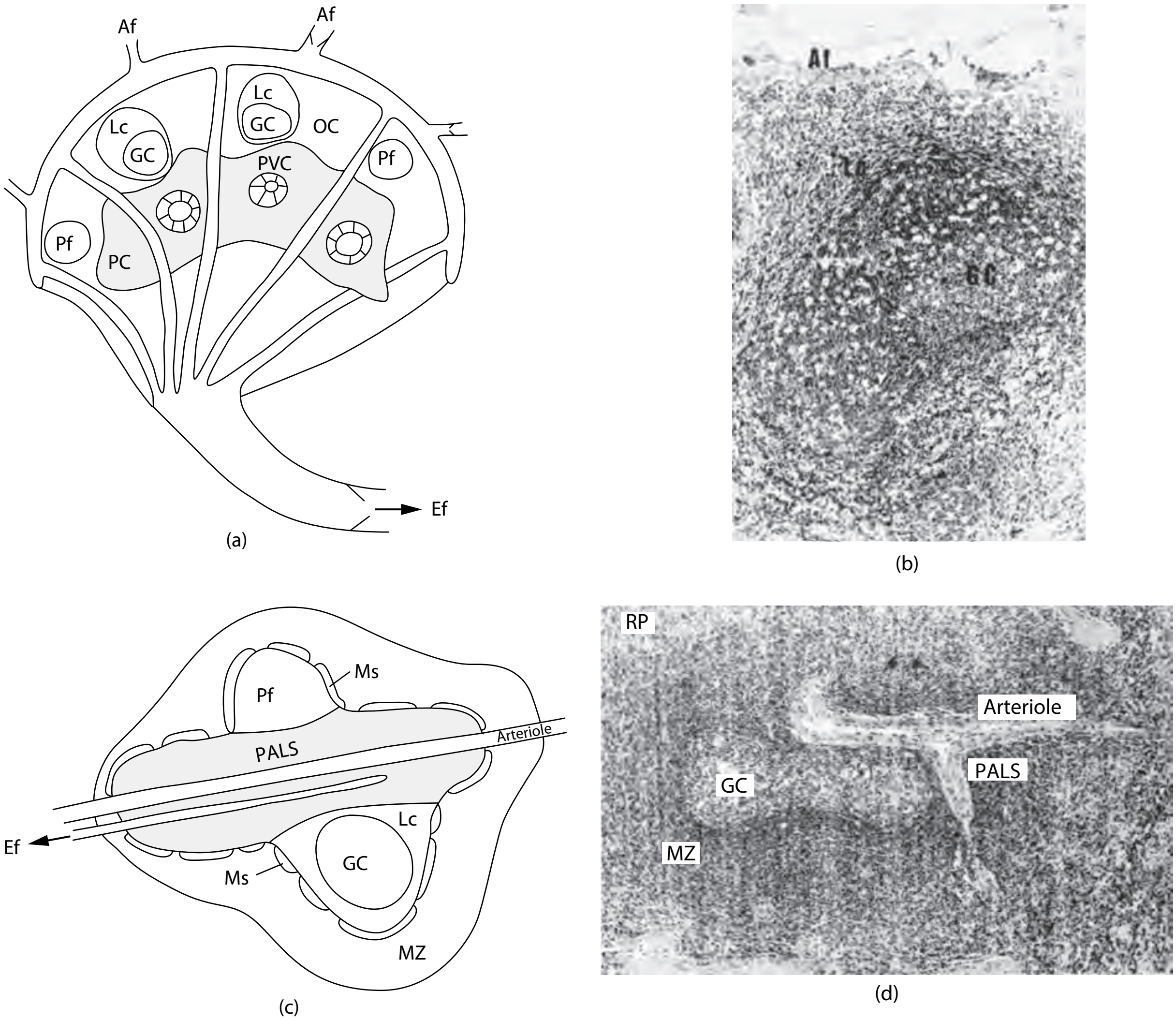


Figure 26.8 Basic arrangement of lymphoid tissue in lymph node and spleen. (a) Diagram of the lymphatic arrangement in a lymph node. (b) Photomicrograph of a section of the spleen showing some of the features illustrated in (a). (c) Lymphatic arrangement in the splenic Malpighian body. (d) Photomicrograph of a section of spleen illustrating some of the features shown in (c). Af, afferent lymphatic; OC, outer cortex; Pf, primary follicle; GC, germinal centre; Lc, small lymphocyte corona; PC, paracortex; PCV, postcapillary venule; Ef, efferent lymphatic; MZ, marginal zone of large lymphocytes; Ms, marginal sinus; PALS periarteriolar small lymphocyte sheath; RP, red pulp.

There is a described phenomenon of 'ineffective lymphopoiesis' within the spleen and lymph nodes. The nuclear remains of pyknotic cells are phagocytosed by macrophages, giving rise to so-called 'tingible bodies'. It may be that these represent 'forbidden clones' (capable of destroying self) and are recognized and destroyed at this stage within the spleen. The spleen may also be the major source for the production of suppressor T-cells. It is further thought that each secondary follicle produces polyclonal 'memory' -cells. Such follicles may also be responding to more than one antigen.

A cellular immune response can be recognized by T-cell enlargement within the paracortex, cytoplasmic RNA synthesis, and cell division. Some cells leave the lymph nodes by way of the efferent lymphatics travelling up the thoracic duct to populate other lymphoid areas via the bloodstream. Efferent lymphatics from the spleen run adjacent to the arterioles, but most lymphoid cells probably leave in vascular channels originating in the white pulp. These channels continue through the marginal zone and open into the red pulp sinuses.

These observations reinforce the fact that lymphoid tissues are not static collections of cells. There is a constant motion of cells through a reticulin scaffolding. Specific stimuli and reactions affect particular components of these populations. When clones of 'well differentiated' cells emerge, as for example, in chronic lymphatic leukaemia, they frequently retain the capacity to migrate, differentiate and form deposits in other distant body sites, e.g. the bone marrow. Structural tumours analogous to follicle centres can arise within a lymph node or group of nodes (follicle cell centre tumours of Luke), and as such may give rise to clones of circulating small cleaved cells that infiltrate and populate other lymphoid areas without causing architectural distortion. Such tumours of small, round, recirculating lymphocytes and follicle centre cleaved cells account for 80% of the acute non-Hodgkin's lymphomas (NHLs) seen in the UK.

Monocytes are sequestered in the white pulp, the marginal sinuses and the red pulp of the spleen. It is in the red pulp where they are converted to fixed macrophages, endowing the spleen with its significant phagocytic capability. It is thought that the spleen, in disease states such as malaria, releases a humoral substance that acts on the bone marrow to cause release of additional monocytes. They eventually circulate to the spleen and can repopulate or further augment its phagocytic capabilities.

Following splenectomy, the primary antibody response is decreased compared to normal, and the secondary response is abnormal in that there is an impaired switching from IgM to IgG antibody subtypes. In addition to antibody synthesis, the spleen produces non-specific effectors of the immune response, such as the tetrapeptide tuftsin. Tuftsin (named after Tufts University where it was identified and characterized) opsonizes particulate matter, and as such facilitates phagocytic activity. It is virtually absent in the blood of asplenic patients. The spleen also influences the opsonization of pneumococci in non-immune individuals and is involved in the alternative pathway of complement activation. All these mechanisms probably account for the increased susceptibility to postsplenectomy sepsis, especially in infants and children. The spleen and lymph nodes both contribute to the immunological surveillance of the host, whereas the spleen additionally performs an important haematological role (Box 26.1).

BOX 26.1 Functions of the spleen and lymph nodes

- Immunological (both spleen and lymph nodes)
  - Antibody production and cell-mediated responses
  - Phagocytosis
  - Maturation of lymphoid cells
  - Significant lymphopoiesis
  - Source of suppressor T-cells
- Haematological (primarily spleen)
  - Filtration of particles from blood: non-specific or antibody coated
  - Removal of red cell inclusions
  - Destruction of senescent or abnormal red cells
  - Compensatory haemopoiesis
  - Storage of platelets, iron and factor VIII

Disorders of the spleen

The most frequently recognized complication of the diseased spleen is the pathological destruction or pooling of blood elements. Splenic enlargement, as it occurs with venous thrombosis and congestion, causes entrapment and pooling that results in destruction of normal cells. This is an exaggeration of the normal function of the macrophage-laden vasculature of the spleen, i.e. the destruction of abnormal red cells, white cells or platelets as they are filtered through the spleen.

Hypersplenism

Hypersplenism is a syndrome of splenomegaly combined with decreased amounts of one or more circulating blood elements:

- anaemia – in patients with diseased bone marrow
- leucopenia $<4–5000/mm^3$
- thrombocytopenia $<100\,000/mm^3$.

Early on, hypersplenism was used to describe the essentially normal function of the spleen in destroying the abnormal cells of hereditary spherocytosis. Damashek broadened this definition to include *all conditions of splenomegaly with decreased numbers of one or more of the blood elements.* Although he postulated that the spleen was active in the control of haemopoiesis via humoral mechanisms, this has no current scientific basis. Some distinguish between primary (due to haematological disease) and secondary (e.g. liver disease with portal hypertension) hypersplenism.

Work hypertrophy of the spleen occurs when the spleen enlarges due to the constant exposure of the spleen's phagocytic machinery to abnormal cells. This is distinctly different from splenomegaly, where destruction of normal blood elements occurs because of a *primary lymphoreticular* process. The term primary hypersplenism is usually reserved for this latter group of disorders as distinct from hyperactivity following splenic enlargement from other causes, e.g. liver disease – *secondary hypersplenism*. In hypersplenism states, the bone marrow is unable to maintain normal numbers of circulating cells or platelets, and splenic enlargement signals the site of destruction. Splenectomy is potentially curative of the cytopenias that occur in hypersplenic states but there are other alternatives (see below).

The exact cause of the cytopenia that occurs with splenic enlargement is a matter of some speculation. The pooling of blood cells is probably the most important mechanism. With increasing splenic size, up to 50% of the total red cell mass may reside in the spleen. This effectively sequesters red cells in a location 'outside' the circulation. The amount of this isolation of red cells can be estimated by measuring peripheral venous haematocrit and comparing that with the red cell volume obtained by isotope dilution techniques.

In a similar fashion, platelets are abnormally sequestered in states of splenomegaly. Approximately 10% of circulating platelets are contained within the spleen in the healthy state. With significant splenomegaly and pooling of blood elements, up to 90% of the circulating platelets are trapped within the spleen. The same can occur with circulating granulocytes and lymphocytes.

The overall effect of such congestive splenomegaly is a dynamic balance between the trapping of blood elements and the ability of the bone marrow to compensate. In disorders with a diminished bone marrow reserve, such as myelosclerosis or chronic myeloid leukaemia, severe anaemia, thrombocytopenia and/or neutropenia may result. If the marrow is healthy or minimally diseased, the peripheral smear may be quite normal.

Hypervolaemia is also a feature of splenomegaly. Where red cell production is limited, the expanded blood volume is mainly plasma volume, with a resultant dilutional anaemia that can compound the destructive anaemia. It is thought that a hyperkinetic portal circulation from an increased splenic blood flow somehow causes an expansion of the splanchnic blood volume. The exact mechanism is far from clear, however. Because of the expanded circulating volume, transfusion in an attempt to restore appropriate numbers of cellular elements can easily cause circulatory overload.

Destruction of blood cells by the spleen

In addition to pooling and trapping within the spleen, the survival of red cells is probably shortened because of 'metabolic stresses'. These include glucose deprivation, lactate accumulation and cellular acidosis due to abnormally close packing of the red cells. The haemolytic component of this specific type of hypersplenism is mild, however. There is less solid evidence for the thrombocytopenia and neutropenia due to these mechanisms.

The spleen routinely destroys abnormal red cells. The cytoskeletal defect manifest as hereditary spherocytosis causes these abnormal red cells to be particularly susceptible to destruction within the spleen. In addition to destroying senescent and diseased red cells, the spleen is active in policing for red cell inclusions such as Heinz bodies (haemoglobin precipitates), via selective membrane disruption and capture of these undesirable intracellular defects.

The spleen also maintains surveillance of surface abnormalities of cells and platelets. In this way, red cells coated with antibody are destroyed via macrophage recognition of the constant portion of the IgG, producing autoimmune haemolytic anaemia. A similar mechanism is probably responsible for the destruction of platelets in certain forms of idiopathic or drug-induced thrombocytopenic states. Additionally, the neutropenia seen in *Felty syndrome* (rheumatoid arthritis and splenomegaly) has an immune basis, probably due to a circulating leucocyte-specific antinuclear factor. The conditions in which splenomegaly can occur with varying degrees of hypersplenism are listed below.

- Infections
  - Acute: hepatitis, mononucleosis, salmonellosis, toxoplasmosis, cytomegalovirus, tularaemia, abscess
  - Subacute: AIDS, bacterial endocarditis, tuberculosis, brucellosis, malaria, leishmaniasis, trypanosomiasis, histoplasmosis
  - Chronic: fungal disease, syphilis, bacterial endocarditis
- Congestive
  - Intrahepatic portal hypertension: cirrhosis, Wilson disease, haemochromatosis, congenital hepatic fibrosis
  - Prehepatic portal hypertension: portal vein thrombosis, obstruction, cavernoma, atresia
  - Posthepatic: Budd–Chiari, congestive cardiac failure
  - Segmental (left-sided portal hypertension): splenic vein occlusion by pancreatitis, pancreatic neoplasm, pancreatic pseudocyst, splenic artery aneurysm
- Haematological
  - Haemolytic disorders: hereditary cell membrane defects, autoimmune haemolytic states (warm antibodies), thalassaemia, sickle cell disease, haemoglobin C disease
  - Myeloproliferative disorders: myeloid metaplasia, polycythaemia vera, essential thrombocythaemia
  - Miscellaneous: primary splenic hyperplasia, megaloblastic anaemia, iron deficiency
- Malignant
  - Haematological malignancies: acute or chronic leukaemias, leukaemic reticuloendotheliosis, malignant lymphomas, malignant histiocytosis, myelomatosis
  - Primary intrinsic malignancies: lymphosarcoma, plasmocytoma, fibrosarcoma, angiosarcoma
  - Secondary malignancies: carcinoma, melanoma
  - Benign: hamartoma, fibroma, haemangioma, lymphangioma
- Inflammatory or granulomatous
  - Felty's syndrome, systemic lupus erythematosus, rheumatic fever, serum sickness, sarcoidosis, beryllosis
- Storage disease
  - Gaucher disease, Wilson disease, Niemann–Pick syndrome, histiocytosis X, Hurler syndrome, Tangier disease
- Miscellaneous
  - Cysts: parasitic, pseudocysts, congenital, traumatic
  - Others: hyperthyroidism, Osler–Weber–Rendu syndrome, splenic mastocytosis, Albers–Schonberg disease.

In practice however 45% of cases of hypersplenism are caused by haematological disease and 40% by liver disease with portal hypertension. Thus all the other causes together account for only 5%.

Management of hypersplenism associated with chronic liver disease

Whereas splenectomy is the recognized treatment for hypersplenism due to haematological disease as it usually normalizes all the components of the cytopenia, the management of hypersplenism associated with chronic liver disease remains non-standardized and currently there are a number of options with no comparative data on efficacy and morbidity. To a large extent, therefore, the treatment depends on local expertise and preferences. Factors that should influence management in the individual case include the severity (Child–Pugh stage) of the disease and the need for treatment of other complications, e.g. bleeding varices, the development of hepatoma in a cirrhotic patient or an end-stage cirrhotic patient requiring liver transplantation.

The most important aspect of hypersplenism in patients with chronic liver disease is the thrombocytopenia, and it is this that requires correction in these patients. Active management is considered only in patients in whom the splenic congestion and the thrombocytopenia are severe. In general, open splenectomy is ill advised in patients with liver disease and portal hypertension because it constitutes a major high-risk operation and may be followed by thrombosis of the portal vein.

The options for the treatment of severe thrombocytopenia in patients with liver disease and portal hypertension are:

- DSRS (Warren procedure)
- transjugular intrahepatic portosystemic shunt (TIPSS)
- laparoscopic splenectomy
- partial splenic embolization (PSE).

DSRS effectively reverses the profound thrombocytopenia resulting from presinusoidal portal hypertension or stable cirrhosis without sacrificing the spleen, and in some centres is considered the treatment of choice for this condition in children. It is used as the only procedure (Child A/B patients) or before hepatic transplantation (Child C disease).

TIPSS improves the hypersplenism in patients with portal hypertension. In one large reported series, the leucopenia was reversed in 50% and thrombocytopenia in 75% of patients. These results indicate that TIPSS is a minimal access intervention that is effective in the treatment of complications of portal hypertension including secondary hypersplenism and has largely replaced the use of DSRS. The problem with TIPSS is its poor long-term patency (beyond 6–12 months). The ideal indication seems to be patients with Child C disease awaiting liver transplantation.

In adult patients with end-stage liver disease and hypersplenism awaiting liver transplantation, and in patients with Child A/B disease requiring hepatic resection for small hepatoma, laparoscopic splenectomy prior to the liver resection/transplantation has been performed as an alternative to TIPSS. Others advise concomitant splenectomy at the time of hepatic resection.

The fourth option in the management of patients with portal hypertension and hypersplenism is PSE. This has to ablate 50% or more of the splenic parenchyma to be effective. It was initially reported to be successful in reversing the hypersplenism (thrombocytopenia and neutropenia) in 90% of patients undergoing embolization of hepatocellular carcinoma and appears to be safe with a low reported incidence of severe complications including splenic abscess formation. PSE was subsequently extended to Child A or Child B (but not Child C) patients with liver disease, where it is reported to also normalize the levels of cholinesterase, total cholesterol and prothrombin time in addition to haematological improvement of the hypersplenism.

Hyposplenism

The more common causes of asplenism and hyposplenism are:

- splenectomy
- splenic agenesis
- atrophy
- coeliac disease
- inflammatory bowel disease and collagenous colitis
- systemic amyloidosis
- alcoholism
- old age
- dermatitis herpetiformis
- sickle cell anaemia
- systemic lupus erythematosus.

The haematological features of hypo/asplenism are:

- abnormal red cells
  - Burr cells
  - target cells
  - pitted cells
- red cell inclusions
  - Howell–Jolly bodies
  - siderotic granules
  - abnormal platelet morphology
  - thrombocytosis
- leucocytosis
  - neutrophilia
  - lymphocytosis
  - monocytosis.

Hyposplenism is confirmed by the appearance of red cell defects (that normally would be culled from the circulation), and the degree of impairment may be qualitatively assessed by ^{99m}Tc-sulphocolloid scintiscanning. There is some evidence based on the percentage of pitted cells in the peripheral blood that splenic function is impaired in elderly people and in alcoholics. Pitted erythrocytes are found in 40% of alcoholics and this functional hyposplenism has been attributed to a direct effect of alcohol on the spleen as the pitted erythrocyte count drops in patients who give up the alcohol habit. However, the possibility that the high pitted erythrocyte count is caused by a direct effect of the alcohol on the red cell membrane can equally account for these changes and the controversy remains unresolved.

The most frequent cause of hyposplenism is surgical splenectomy. Congenital asplenia (*Ivemark syndrome*) is a truly rare disorder, associated with complex cardiac, gastrointestinal, genitourinary and neuromuscular abnormalities. Survival in such cases is largely determined by the ability to recognize and correct the complex cardiac defects. Gastrointestinal anomalies such as malrotation and situs inversus are frequently seen in congenital asplenia. In surviving infants, a common cause of death is overwhelming sepsis due to encapsulated organisms, particularly pneumococcal.

Splenic hypoplasia can also occur from birth as part of the syndrome of *Fanconi's anaemia*, i.e. congenital hypoplastic anaemia. Acquired hyposplenism occurs in about 76% of patients with coeliac disease. The cause is unknown, but may be related to the increased absorption of dietary antigen and subsequent overload of the spleen with circulating immune complexes. The hyposplenic state improves with a gluten-free diet in these individuals. A relationship has also been noted between the morphological state of the intestinal epithelium and splenic function. Hyposplenic states have also been described in other disorders involving the gastrointestinal mucosa including Crohn disease, ulcerative colitis, collagenous colitis and intestinal lymphangiectasia.

Circulating autoantibodies and immune complexes in clinical autoimmune disorders, e.g. systemic lupus erythematosus, have been noted to cause a functional hypoplastic state secondary to Fc-receptor blockage. The hyposplenism of sickle cell anaemia is related to the degree of splenic infarction. Hyposplenism is also a feature of systemic amyloidosis. Hyposplenism can also occur in patients with full-blown HIV and these patients usually present with the *Mycobacterium avium* complex (infection) complicating HIV-related immune thrombocytopenic purpura. This combination is usually lethal, as the severe thrombocytopenia does not improve with corticosteroids, intravenous immunoglobulin and splenectomy. The peripheral blood smear shows Howell–Jolly bodies. Pathological examination of the spleen shows multiple granulomas with numerous acid-fast organisms.

Postsplenectomy sepsis

There is now uniform consensus that the risk of overwhelming sepsis is increased significantly after splenectomy. There is good evidence that the greatest risk is in infants and children (first 5 years of life), and to some extent the risk in adults may have been overestimated in the past, although this is debatable. The problem of postsplenectomy sepsis has been compounded by the increasing incidence of penicillin-resistant pneumococci. The risk of postsplenectomy sepsis is also influenced by the nature of the disease for which the spleen is removed, with the lowest incidence of sepsis being reported after splenectomy for trauma as these individuals are otherwise healthy and have no other disease that impairs the immune defence against invading bacteria. The reported incidence of postsplenectomy sepsis in relation to indication for splenectomy is shown in Table 26.2.

Table 26.2 Incidence of postsplenectomy sepsis in relation to indication for splenectomy

Reason for splenectomy	Incidence of sepsis (%)
Trauma	1.4
Immune (idiopathic) thrombocytopenia	2.0
Incidental (iatrogenic injury)	2.1
Congenital spherocytosis	3.5
Acquired haemolytic anaemia	7.5
Portal hypertension	8.2
Primary anaemia	8.5
Reticulosis/lymphomas	11.5
Thalassaemia	24.8

Modified from Singer, *Persp Pediatr Pathol* 1973;**1**:285–311.

Although the absolute lymphocyte count is increased after splenectomy for trauma, the peripheral blood lymphocyte subpopulations of these otherwise healthy subjects are distinctly abnormal. There is a significant reduction in the percentage of CD4$^+$ T-cells due to a selective and long-term decrease in the percentage of CD4$^+$CD45RA$^+$ lymphocytes. These decreased levels of CD4$^+$CD45RA cells are accompanied by an impairment of the primary immune responsiveness in terms of both T-cell proliferation and antibody responses to newly encountered antigens. By contrast, levels of the reciprocal CD45RO$^+$CD4$^+$ T-cell subset, lymphoproliferative responses and interferon γ production to recall antigens remain normal. These findings suggest that the intact spleen is essential for the generation, maintenance and/or differentiation of unprimed T-cells or their precursors and may explain the impaired primary immune responses following splenectomy.

Splenectomy for trauma carries the lowest risk and thalassaemia the highest; but even for the lowest risk group, there is still a 40- to 50-fold increase in the incidence of overwhelming sepsis. Some of these estimates have been questioned largely because the majority of reported series have had small numbers, and indeed a substantial cohort of the data (on which estimates are based) are from single case reports of pneumococcal serious infections, often with bacteraemia. It is an undeniable fact that community-acquired pneumococcal pneumonia with bacteraemia is common in patients with normal splenic function and is seldom reported because of its established occurrence in susceptible groups. Because of this, there are some who argue that the risk of postsplenectomy pneumococcal sepsis may have been exaggerated.

Although *Streptococcus pneumoniae* is implicated in over 55–60% of septic episodes in asplenic patients, infections by other encapsulated bacteria, e.g. *Haemophilus influenzae, Haemophilus pertussis* and *Neisseria meningitidis*, are also more commonly reported in these individuals, as indeed are infections by Gram-negative bacteria, e.g. *Escherichia coli*. The functional deficits in asplenic patients are numerous, but impaired filtration, diminished phagocytosis, decreased IgM levels and loss of the opsonic tetrapeptide tuftsin all contribute to the increased risk.

The syndrome of overwhelming postsplenectomy infection (OPSI) often begins insidiously. From a non-specific viral-like illness or malaise, the course rapidly turns fulminant. High fever, nausea, vomiting, dehydration, hypotension and obtundation occur with alarming speed, death being the end result if the course of events is not quickly reversed by effective resuscitation and aggressive antibiotic therapy. Gram stain of the peripheral blood smears will occasionally reveal the causative organism and blood culture is invariably positive. The mortality rate of OPSI is 50–80% for all cases combined. Postmortem examination often reveals bilateral adrenal haemorrhage (*Waterhouse–Friedrichsen syndrome*).

The treatment of suspected episodes of OPSI should be aggressive and without any delay. Broad-spectrum antibiotics effective against encapsulated cocci in the first instance are administered intravenously. Intravenous colloids are used to correct the hypovolaemia using CVP and urine output monitoring within an intensive care setting. Altered haemostasis and disseminated intravascular coagulation can occur, and replacement products (fresh frozen plasma, platelets) are often required.

The prevention of OPSI is imperfect at best and is based on vaccination, administration of oral penicillins and patient education. Prophylactic antibiotic therapy is recommended (together with vaccination) in children and some would extend it indefinitely to adults. Although this is sensible for the high-risk groups, there is less of an argument for long-term prophylactic antibiotics in adults undergoing splenectomy for trauma or iatrogenic injury sustained during abdominal surgery. The vaccination should be carried out at least 10–14 days prior to splenectomy for maximum effective immunization. The vaccination programme should always include the polyvalent pneumococcal vaccine and *Haemophilus* vaccines. Unfortunately, the pneumococcal vaccine (against 23 of the most prevalent strains) does not provide immunity against all strains. Not all vaccinated patients develop immunity but the majority do, and the titres of antipneumococcal antibodies remain elevated for up to 42 months. The question of booster immunizations has not been resolved. Immunization after splenectomy (trauma cases) is much less effective but some advise it anyway.

Since no form of prophylaxis is completely effective, it is very important that all patients carry 'splenectomy' cards, and close surveillance and specific patient education are mandatory. At the first sign of infection, all patients with hyposplenic states should be strongly advised to seek medical attention and receive antibiotic therapy.

Splenic infarction

Infarction of the spleen is not uncommon and its clinical presentation is variable. Thus some patients have severe acute symptoms and signs, and may develop serious life-threatening complications (splenic rupture, splenic abscess), whereas others (up to 30%) have minor symptoms or are asymptomatic.

Aside from the predictable autoinfarction in patients with sickle-cell disease, splenic infarction occurs most commonly with splenomegaly due to congestive disorders such as chronic myeloid leukaemia and myelosclerosis. Thromboembolic disorders may cause splenic infarction, e.g. atrial fibrillation, arterial embolic (atheromatous) disease, diabetes-associated microvascular disease and acute torsion of an *ectopic* (*wandering*) *spleen*. A variety of other disorders may be complicated with the development of splenic infarction, e.g. falciparum malaria, Q fever, AIDS, severe necrotizing pancreatitis and pancreatic pseudocysts. Splenic infarction has also been reported as a complication of injection of gastric varices with Histoacryl. The splenic infarction in patients with AIDS is associated with the development of high titres of anticardiolipin antibodies, thrombocytopenia and a coagulopathy. Episodes of cerebrovascular infarction may also occur in these patients. The cause of the splenic infarction in these AIDS patients is arterial thrombosis of the coeliac trunk.

The age range of splenic infarction varies widely (children to old age) and 70% are symptomatic. The most common symptoms are acute upper left quadrant abdominal pain, fever, chills and malaise. However, some patients are asymptomatic and these are usually patients with non-malignant haematological disorders. Fever and leucocytosis are especially marked in patients with thromboembolism as the cause for the splenic infarction. Physical examination reveals tenderness and guarding maximal in the left upper abdomen. Splenic infarction results in a capsular inflammatory reaction frequently causing irritation of the left diaphragm. This may result in left basal pleurisy/effusion and pain referred to the left shoulder (Kerr's sign). Confirmation of the diagnosis is usually achieved by abdominal angio-CT scan.

Initially, the management should be conservative with analgesia and antibiotics. Surgery is indicated if the diagnosis is in doubt or for complications (splenic abscess, bleeding from splenic rupture) when splenectomy is indicated. The morbidity (mainly pulmonary complications) is high and the mortality averages 5% overall.

Pathological changes in spleen and lymph nodes

The pathological processes affecting the spleen and/or lymph nodes can be categorized as:

- reactive
- NHLs
- Hodgkin disease
- secondary tumour deposits.

Immunological reactivity

Reactive changes cause regional lymphadenopathy and, depending on the cause, may be non-specific, granulomatous, epithelioid congeries and angio-immunoblastic with or without accompanying changes in the spleen. These reactive changes are categorized in Table 26.3.

The spleen and lymph nodes normally respond to antigen loads with an expansion of the mononuclear–macrophage and lymphocyte cell lines specific to the challenge. An antibody response results in the increased numbers of immunoglobulin-secreting -cell proliferations in the medullary cords of the lymph nodes or red pulp of the spleen. Reactive follicle centres develop secondarily. Although bacteria that are coated with antibody are more rapidly cleared by the macrophage–monocyte/neutrophil system than non-coated cells, it may be that such coating is counterproductive in the surveillance against tumour cells. Antibody-coated tumour cells may be less easily recognized as 'foreign' by cytotoxic T-cells.

T-cell proliferation follows vaccination or cellular challenges, and the result is expansion of the paracortex surrounding the epithelioid venules. This situation can simulate neoplasia even to the extent of producing binucleate, large cells resembling Reed–Sternberg cells. However, in this situation the architecture of the spleen or lymph node is maintained, and secondary follicle reactivation with plasma cells is usually seen.

Non-antigenic material within lymphatics, such as carbon pigment, is trapped in the medullary sinuses of lymph nodes. Bloodborne material, such as haemosiderin, is trapped by splenic red pulp macrophages. Such non-antigenic entrapment can cause nodal enlargement, e.g. dermatopathic lymphadenopathy, Gaucher disease (lipid storage disease) or thalassaemia/haemolysis. Expansion of the histiocytic elements of lymph

Table 26.3 Reactive pathological changes caused by disease of lymph nodes and spleen

Reactive	Lymph nodes	Spleen
Non-specific		
Sinus histiocytosis	Expansion of medullary cords by histiocytes Removal of non-antigenic material (carbon, Hb) or fat (dermatopathic lymphadenopathy)	Increased numbers of macrophages in red pulp
Follicle centre reactivity	Expanded reactive follicle centres. Numerous mitoses and tingible body macrophages	Reactive secondary follicles in Malpighian bodies
Paracortical	Expansion of paracortex by transforming T-cells, e.g. viral infections	Expansion of periarterial lymphoid sheath by transforming T-cells
Plasma cell	Large numbers of polyclonal immunoglobulin-containing plasma cells in medullary cords	Large numbers of plasma cells in red pulp
Granulomatous		
Non-caseating	Cortical epithelioid granulomas with Langerhans giant cells, reticulin production with no necrosis – sarcoid reaction, e.g. draining tumours, beryllium poisoning	Similar granulomas Initially in the marginal zones
Caseating	Central caseous necrosis, with fibrosis, e.g. tuberculosis	May show amyloid deposition
Necrotizing	Stellate microabscesses with polymorphs and peripheral histiocytes, and multinucleated giant cells, e.g. cat scratch fever, lymphogranuloma venereum	
Epithelioid congeries	Collections of small numbers of epithelioid histiocytes without caseation or giant cells in paracortex with active secondary centres, e.g. toxoplasmosis, visceral leishmaniasis (basophilic Leishman - Donovon bodies)	Red pulp greatly expanded by histiocytes containing Leishman - Donovon bodies in leishmaniasis
Angioimmunoblastic lymphadenopathy	Reactive hyperimmune state with proliferation of epithelioid venules, interstices filled with transforming lymphoid cells, polyclonal immunoblasts and mature plasma cells. May burn out leaving sclerotic nodes or progress to malignant immunoblastic sarcoma	

nodes and spleen occurs with parasitaemia (leishmaniasis, malaria) and to a lesser extent with toxoplasmosis and systemic leprosy. Malarial infestation adds immunoproliferative responses to such histiocytic expansion.

Lymphomas

Lymphomas are the common primary tumours and are classified as Hodgkin's and non-Hodgkin's. Secondary tumours (deposits) affect largely the lymph nodes, as, although they may involve the spleen, splenic metastases are very rare. The spectrum of NHLs arises in cells that typically are migratory throughout the lymphatic system. As such, the more differentiated a tumour, the more difficult it may be to demonstrate malignancy before gross architectural distortion occurs. This spectrum is evident by describing two diseases: *chronic lymphatic leukaemia* where the overproduced small B-lymphocytes continue to circulate, and *lymphocytic lymphoma* where a small population remains confined to lymph nodes. This wide variety of pathological states certainly makes it difficult for the pathologist and clinician both diagnostically as well as in recommending therapeutic action.

It is possible to demonstrate a clone of abnormal cells within the circulating pool of lymphocytes. Thus peripheral blood can be diagnostic when architectural distortion of the lymph nodes or spleen is not diagnostic. This also serves to confirm that, even in well-differentiated lymphomas' the disease is usually systemic, and not just confined to the enlarged lymph nodes/organs.

Intermediate and high-grade lymphomas more typically destroy locally and invade, rather than migrate widely. They are also often highly sensitive to radiotherapy and chemotherapy. As they can grow quickly, disseminate late and are often noticed early by the patient, local therapy alone may be effective. It is because of this wide spectrum of presentation and variable systemic involvement that accurate clinical staging is imperative in the choice of therapeutic regimens. This is especially true, paradoxically, for the cytologically more 'benign' lower grade tumours, which are often metastatic at the time of initial diagnosis. The histopathology/classification of Hodgkin disease according to the revised European–American system is as follows:

- I–lymphocyte predominance
- II–provisional entity: lymphocyte rich–classical Hodgkin disease
- III–mixed cellularity
- IV–nodular sclerosis
- V–lymphocyte depletion.

It is important to stress that, in Hodgkin disease, a portion only of the lymph node may be involved and typically the spleen shows scattered foci of the disease. This is in sharp contrast to NHLs where there is total involvement of the nodes or the spleen. The classification of NHL according to the revised European classification of NHLs recognizes two main groups (B- and T-cell lymphomas) and several other rarer types.

-cell lymphomas

β-cell neoplasms: precursor B-lymphoblastic leukaemia/lymphoma.

- Peripheral -cell lymphomas

1 -cell chronic lymphocytic precursor leukaemia/small lymphocytic lymphoma

2 Lymphoplasmatoid lymphoma/immunocytoma
3 Mantle cell lymphoma
4 Follicle centre lymphoma
5 Marginal; zone -cell lymphomas extranodal (mucosa-associated lymphoid tissue (MALT) type monocytoid -cells) Subtype: nodal (monocytoid -cells)
6 Splenic marginal zone lymphoma (villous lymphocytes)
7 Hairy cell leukaemia
8 Plasmacytoma
9 -cell large lymphoma: subtype: primary mediastinal (thymic) -cell lymphoma
10 Burkitt's lymphoma
11 High-grade -cell lymphoma, Burkitt's-like

T-cell and natural killer (NK)-cell lymphomas

Precursor β-cell neoplasms: precursor T-lymphoblastic lymphoma/leukaemia.

- Peripheral T-cell and natural killer (NK) cell neoplasms

1 T-cell chronic lymphocytic leukaemia, prolymphocytic leukaemia
2 Large granular cell lymphocytic leukaemia
3 T-cell

- Subtype NK-cell type

4 Mycosis fungoides/Sezary syndrome
5 Peripheral T-cell lymphoma, unspecified
  - Subtype: hepatosplenic T-cell lymphoma
  - Subtype: subcutaneous T-cell lymphoma
6 Angioimmunoblastic lymphoma (AILD)
7 Intestinal T-cell lymphoma (enteropathy-associated)
8 Adult T-cell lymphoma/leukaemia
9 Anaplastic large-cell lymphoma, CD30, T- and null-cell types
10 Anaplastic large-cell lymphoma, Hodgkin's like

A special type of extranodal NHL arises from the MALT found in the respiratory epithelium, salivary glands and the gut. This may give rise to MALT-lymphomas or MALTomas, most commonly of the stomach (see Chapter 23).

Hodgkin disease spreads less often along contiguous groups of lymph nodes than NHLs. The disease initially appears as a small focus within an otherwise reactive lymph node (Figure 26.9), and although it metastasizes widely like a carcinoma, actual invasion of the lymphatics rarely occurs.

The more differentiated lymphocyte-predominant variety has a better prognosis than (in decreasing order) nodular sclerosing, mixed cellular or lymphocyte-depleted Hodgkin's. With the routine use of more aggressive radiotherapeutic and chemotherapeutic regimens, the exact cellular classification has become less important in the overall scheme of treatment.

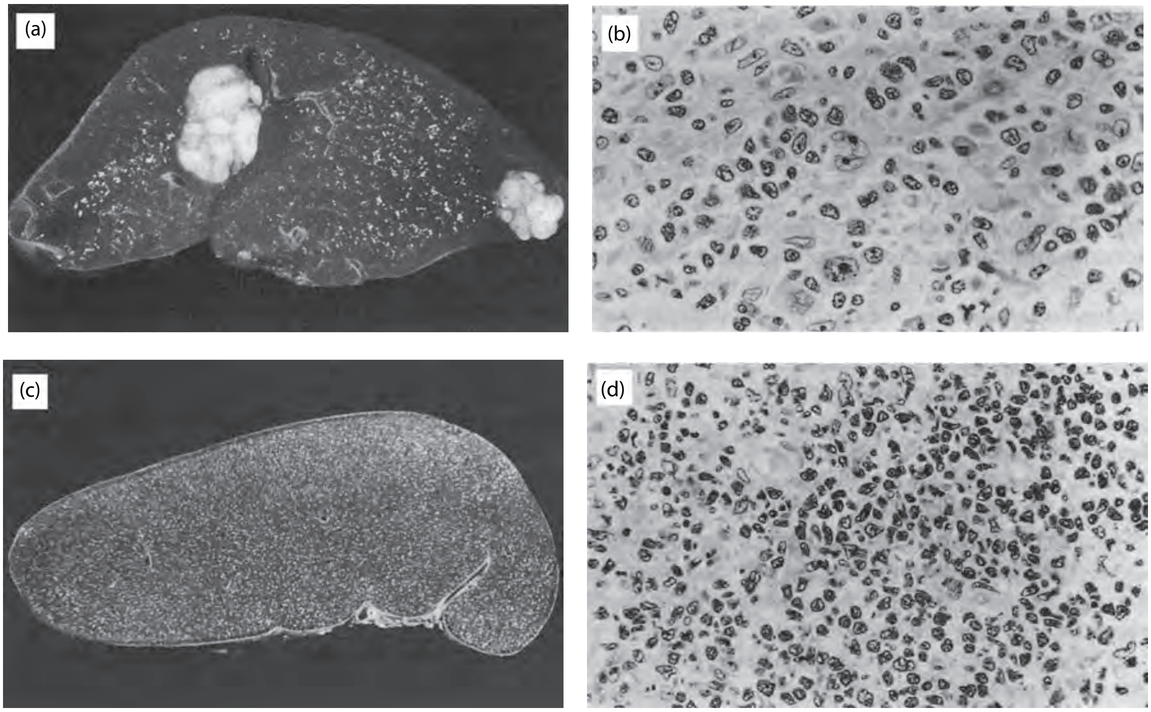


Figure 26.9 The spleen in Hodgkin disease and non-Hodgkin's lymphoma: (a) transverse section of a 200 g spleen removed at laparotomy for staging Hodgkin disease. Two discrete tumour foci are present and histology of the intervening tissue shows only non-specific reactive features. (b) Histology of one of the tumour nodules showing mixed cellular Hodgkin disease, note the central diagnostic 'mirror image' binucleated Reed–Sternberg cell, eosinophils (with bilobed nuclei and granules in the cytoplasm, e.g. three cells from the R-S cell at one o'clock) and plasma cells (e.g. three cells away at three o'clock); the large cells with stippled nuclei are histiocytes. (c) Transverse section of a 400 g spleen from a patient with non-Hodgkin's lymphoma, note involvement of every Malpighian body, histologically there was minimal disturbance of the architecture. (d) Tumour cells from the white pulp of the same spleen showing small cleaved follicle centre cells. This patient had a small number of similar cleaved cells circulating with his peripheral lymphocytes in the blood.

Secondary neoplasms, e.g. myeloid leukaemia, are a particular risk, however, when radiotherapy and alkylating chemotherapeutic regimens are combined.

Familial erythrophagocytosis (of Claireaux and Farquhar) involves sinus histiocytes and splenic macrophages that abnormally engulf and destroy red cells. A severe anaemia occurs, but there is little, if any, architectural destruction of the involved organ. The end stage of this disease evolves into a monocytic leukaemia. A similar non-selective phagocytosis is seen in *histiocytic medullary reticulosis.* While centred in lymph nodes and/or the spleen, the group of *histiocytosis X* diseases is probably not a primary disease of these organs.

Primary haematological disorders

Bloodborne 'haemic reticuloses' are less often found in biopsies of enlarged lymph nodes since such diagnoses are usually first made by other methods. Myeloid leukaemias may be associated with medullary lymph node proliferation or splenic red pulp enlargement and the diagnosis is readily made with stains specific to eosinophil granules (e.g. azo-eosin) or histochemical staining for chloracetate esterase (monocytes). In myelosclerosis, these same areas show a myeloid transformation with a high proportion of multinucleated giant cells, similar to those seen in bone marrow or to megakaryocytes. Erythroid cells, contrary to the lymphocyte lines, have a characteristically dense nucleus, so-called 'empty' cytoplasm, and a periodic acid–Schiff-positive spherical cell membrane. Both familial and acquired haemolytic anaemias may produce myeloid metaplasia of the spleen and lymph nodes.

Splenomegaly

As previously outlined, there are numerous causes of splenomegaly and its relative incidence varies in different parts of the world, but in Western hospital practice, the distribution is as follows:

- patients with hepatic diseases, most commonly cirrhosis – 36%
- patients with haematological disease – 35%
- patients with infectious diseases – 16%, increasingly AIDS
- patients with inflammatory non-infectious disease – 5%
- patients with primary splenic disease – 4%
- others – 3%, e.g. congestive heart failure, endocarditis.

During the past 20 years, AIDS has come to account for 55–60% of patients with splenomegaly caused by infectious disease. The spleen may enlarge transiently in a variety of acute bacterial and viral infections, chronic infections and in subacute bacterial endocarditis. Parasitic infections, e.g. malaria, can result in massive congestive splenomegaly such that rupture is a very real and well-documented risk in those affected.

Portal hypertension causes mild to moderate splenomegaly unless it is posthepatic (Budd–Chiari) or follows splenic vein thrombosis (sectorial, left-sided, sinistral portal hypertension) when massive splenomegaly may occur.

Splenomegaly accompanies both hereditary and acquired red cell defects. The increasing splenic size predisposes these patients to increased red cell surveillance and hence destruction on a volume basis alone. As such, a vicious cycle is established whereby increased destruction and splenomegaly produces even more rapid red cell destruction.

Splenomegaly occurs in about one-third of patients with megaloblastic anaemia, but less frequently in iron-deficiency anaemia. Splenomegaly is unusual in acquired aplastic anaemias, although for unexplained reasons it occurs in a significant number of children with congenital hypoplastic anaemia. It also occurs quite frequently in patients with dyserythropoietic anaemia or 'preleukaemic' states, e.g. sideroblastic anaemia.

Splenomegaly regularly accompanies myeloproliferative disorders. Although polycythaemia vera does not commonly exhibit extramedullary erythropoiesis, the process of myeloid metaplasia in myelosclerosis does and splenomegaly is often massive in this condition. Any form of leukaemia or lymphoma can also cause splenomegaly via the same mechanisms. As such, the splenomegaly that accompanies Hodgkin disease or NHL does not necessarily represent tumour involvement of the organ, raising the possibility of a false-positive clinical stage.

Clinical features

A left upper abdominal mass is often not splenomegaly. For this reason, a comprehensive history and good physical examination is necessary in the first instance to minimize the considerable expense and discomfort of a misdirected evaluation. The normal-sized spleen is not palpable but, when palpable, the spleen is at least two to three times the normal size. The important questions for history taking reflect the common causes of splenomegaly for the particular region. With the spectrum of diseases in mind, it is important to consider infectious (AIDS in particular) and neoplastic disease. A history of travel is often missed unless specifically questioned. Any suggestion of prior pancreatitis or abdominal pain with alcoholism should raise the suspicion of splenic vein thrombosis. The review of the systems enquiry should include questioning for pruritus, as this frequently accompanies myeloproliferative disorders.

The symptoms of splenomegaly itself (irrespective of cause) do not correlate with size, but in practice they tend to be marked only when the spleen becomes massively enlarged since most are due largely to mechanical displacement of adjacent organs and weight of the congested spleen. The symptoms of splenomegaly, as distinct from the manifestations of the underlying disease, include chronic dragging abdominal pain, or pain when lying on the side, abdominal discomfort and early satiety. In addition patients may complain of attacks of acute (colicky) left upper quadrant pains. In general, the bigger the spleen the worse the cytopenia (from hypersplenism) although there are exceptions to this, e.g. marked thrombocytopenia/neutropenia and immune (idiopathic) thrombocytopenia purpura where the splenomegaly is invariably mild. Thus the significant correlation between splenic size and blood counts is not always clinically significant.

The physical examination of a left upper quadrant mass starts with the examiner's hand well inferior in the right iliac fossa probing gently through each exhaled breath of the patient. In addition, the examiner's left hand is cupped posteriorly on the patient's back and flank so as to produce a 'bimanual' feel to the mass in question. If the mass is an enlarged spleen, it will be impossible to appreciate its superior limit. In contrast to the colon and stomach, the spleen lying against the abdominal wall is dull to percussion. Capsular inflammation of the spleen may produce a rub (with the stethoscope). Finally, the right lateral decubitus position may allow for easier examination of a left upper quadrant mass.

A renal mass can usually be distinguished from a splenic mass on physical examination since:

- the kidney moves inferiorly on respiration, whereas the spleen moves medially as well
- the organ shapes are quite different
- the colonic resonance in front of a renal mass is usually not present with splenic masses.

Colonic and gastric masses usually move less well with respiration as they are not as intimately attached to the diaphragm, are more irregular than splenic masses and often have a superior limit that is palpable. The physical examination must also include a careful search for lymphadenopathy, including the tissues of the posterior pharynx. A search for the stigmata of chronic liver disease as well as evidence of purpura or bruising mandates a complete head-to-toe examination.

Clinically splenomegaly depending on extent of enlargement may be:

- mild – up to 5 cm below the left costal margin
- moderate – enlargement up to the umbilicus
- marked (massive)–more than 8 cm or below the umbilicus.

Massive splenomegaly is defined as a spleen having a craniocaudal length greater than 18–20 cm or weighing more than 600 g. Sometimes the term supramassive splenomegaly is used for spleens with a craniocaudal length greater than 22 cm or weight more than 1600 g but the clinical relevance of this category is moot.

Investigation of splenomegaly

The cause of splenic enlargement can usually be identified by history, physical examination and a few appropriate tests. Those cases associated with haematological disorders are often fully characterized by a peripheral blood smear and a bone marrow biopsy. Lymphoreticular malignancies are defined by appropriate biopsy. Haematological or serological testing can identify most infectious causes of splenomegaly. Mononucleosis, a frequent cause, is diagnosed by the finding of atypical lymphocytes in the peripheral blood smear, a positive Paul–Bunnell test (or similar screen), and a rising anti-Epstein–Barr virus titre. Most patients with AIDS who develop splenomegaly are already known to have the disease, but when suspicion is aroused in previously undiagnosed patients, permission for the appropriate testing should be obtained. Patients with a history of travel or those living in endemic areas of disease should have blood smears looking for malaria or marrow for Leishman–Donovon bodies. Disseminated tuberculosis must be considered in all members of a community where there are also immigrants at risk. Laparoscopy may assist in this elusive diagnosis as the majority will have intestinal tuberculosis.

The size, shape and consistency of the spleen can be accurately visualized by CT, MRI or ultrasonography. The determination of splenic size is of crucial importance in determining the surgical approach when splenectomy is indicated in the management of the patient (see below). The spleen's blood vessels can be imaged using either duplex ultrasound, high-dose contrast helical CT and MRI or (rarely nowadays) with selective visceral angiography. The last is used almost exclusively for therapeutic embolization. It is no longer necessary to puncture the spleen directly for splenoportography, as other methods are usually sufficiently accurate. Fine ultrasound-guided needle splenic biopsy is carried out for specific lesions but only in specialized centres.

Ultrasound is the most widely used imaging modality in the investigation of patients with both acute and chronic disorders of the spleen. It is performed by scanning through the intercostal spaces with both greyscale and colour Doppler or power flow (splenic vasculature). Splenic ultrasound is useful for:

- detection of accessory spleens
- confirmation of splenomegaly but not the cause of the splenomegaly
- differentiation of solid from cystic intrasplenic focal masses
- detection of calcification, wall thickening, internal debris, and gas within cystic lesions
- detection of splenic cavernous haemangiomas
- diagnosis of splenic infarction
- diagnosis of splenic trauma and monitoring patients with splenic injuries managed conservatively.

An accurate assessment of the spleen's function can be obtained through injection of labelled platelets, cells (red, white) or carrier molecules and radiotracer studies. The splenic uptake rate of Tc–sulphur colloid or Tc–tin colloid provides a sensitive and quantitative function of splenic function and is based on tracer uptake by the spleen (measured splenic uptake rate divided by measured injected activity). The normal splenic tracer uptake rate is 0.0002–0.0006/s. Values lower than 0.0002/s indicate hyposplenism and values greater than 0.0006/s hypersplenism. There is good correlation between high splenic tracer uptake rates and the severity of the neutropenia and thrombocytopenia.

Undiagnosed splenomegaly

Patients are infrequently encountered where, despite all reasonable efforts, the cause of the splenomegaly remains cryptic. In an otherwise healthy person, this can be a vexing situation. When the organ is minimally enlarged and the patient otherwise healthy, the management consists of careful follow-up. If symptomatic, noticeably enlarging or already significantly enlarged at presentation (the more common situation), diagnostic splenectomy should be considered. In these patients, even if no specific pathological entity is found, one should continue careful surveillance as a lymphoma may develop years after such a splenectomy (possibly due to sampling error at pathological review). Nonetheless, many of the patients who require splenectomy will be found to have an occult lymphoma.

Splenectomy

Indications for splenectomy

There are definite, desirable and debatable indications for splenectomy.

- Definite indications
  - non-salvageable spleen injury (see text)
  - *en bloc* resection of adjacent neoplasms (usually proximal gastric cancer)
  - neoplasms of the spleen – usually lymphomas
  - splenic abscess
  - echinococcal cysts
  - bleeding gastric varices due to sinistral portal hypertension
  - rupture of diseased spleen
- Desirable indications– used selectively
  - hereditary spherocytosis
  - immune (idiopathic) thrombocytopenic purpura
  - AIDS-related thrombocytopenic purpura
  - autoimmune haemolytic anaemia
  - sickling syndromes (sickle cell disease and sickle-β-thalassaemia)
- Debatable indications
  - non-parasitic splenic cysts
  - thalassaemia syndromes
  - lymphoma with specific cytopenia or pancytopenia
  - Thrombotic thrombocytopenic purpura
  - myeloproliferative disorders.

Definite indications

Neoplasms of the spleen should be removed for accurate diagnosis and staging. Septic emboli to the spleen do not require splenectomy, but, when an abscess has formed, removal of the entire organ is the safest management course. Potential spillage of echinococcal cyst contents requires that the entire spleen be removed for this condition.

Bleeding gastric varices resulting from splenic vein thrombosis require splenectomy as does rupture of a splenic artery aneurysm that can occur catastrophically during pregnancy. Splenectomy may be required (*de nécessité*) during resection for proximal gastric cancer to obtain the necessary clearance and adequate regional lymphadenectomy. Although spontaneous rupture of the spleen does occur, it is most frequently seen at times of, or following an interval after, minor abdominal trauma; or as the presenting symptom of previously silent splenomegaly in an already diseased spleen, e.g. malaria, mononucleosis, myelosclerosis, etc. Rarely, a rapidly enlarging spleen in the aggressive forms of NHL may rupture spontaneously. Equally rare, a pseudocyst of the tail of the pancreas may cause splenic vein thrombosis, rapid enlargement and congestion of the organ and spontaneous rupture. In otherwise healthy patients, rupture of the spleen much more commonly follows blunt trauma of the chest and abdomen (see section Trauma) or as an iatrogenic injury during abdominal surgery. Because of concerns regarding OPSIs, every reasonable effort should be made to salvage an injured spleen during surgery and emergency laparotomy (splenic suture, partial splenectomy, etc.). However, there are situations when splenic salvage is ill advised and splenectomy is the sensible option for:

- hilar injuries or a shattered spleen (grade 4 or 5 injuries)
- blast injuries to the organs of the left upper quadrant
- multiple associated injuries where splenic salvage may prolong the procedure
- haemodynamically unstable and elderly patients
- marked intra-abdominal contamination
- rupture of a pathological spleen.

Non-operative conservative management of a known splenic injury, especially in the paediatric age group, is certainly an acceptable alternative to operative therapy in the first instance. Likewise, in an effort to avoid the many complications of splenectomy in patients with myelosclerosis, a small perisplenic haematoma may be closely observed in hospital, with frequent serial ultrasound examinations. But in either case, with any sign of haemodynamic instability or continued bleeding (need for persistent blood replacement), laparotomy is clearly indicated. The disadvantage of conservative management is that, in the event of failure, the potential for splenic salvage is reduced.

Desirable indications

In many of these patients, the decision regarding the need for splenectomy is made by the haematologist who refers the patient to the surgeon. It is usually desirable to perform splenectomy for hereditary spherocytosis, refractory immune (idiopathic) thrombocytopenic purpura and AIDS-related thrombocytopenia that have failed to respond to medical therapy (steroids and human IgG), and acquired haemolytic anaemia. Some cases of genetic red cell enzyme defects, such as pyruvate kinase deficiency, respond favourably to splenectomy. Splenectomy may reduce the blood transfusion requirements of haemoglobinopathies, such as the thalassaemia syndromes. It may also be useful in patients with neutropenia secondary to congestive splenomegaly, e.g. tropical or lymphomatous, for the same reason as in those with storage diseases.

In patients (usually children) with sickling disorders (sickle cell disease and sickle-β-thalassaemia) splenectomy is beneficial in the management of patients who develop large spleens with hypersplenism, major splenic sequestration crisis, recurrent minor splenic sequestration crises, splenic abscess and massive splenic infarction. A high proportion (25%) of these patients has concomitant gallstones and cholecystectomy may be considered at the time of the splenectomy.

Debatable/controversial indications

Splenectomy in the treatment of myeloproliferative disorders is controversial. Its value in the management of chronic myeloid leukaemia remains unproven. It may be occasionally indicated in myelosclerosis with massive splenomegaly, but only if the sequestration of red cells exceeds the spleen's erythropoiesis as measured by isotope tests. The 'rebound thrombocytosis' often seen after splenectomy is particularly severe in this situation. As myelosclerosis follows an unpredictable course of severity, splenectomy to avoid these complications is highly debated. Staging laparotomy for Hodgkin's and NHLs with splenectomy, lymph node harvest and liver biopsies is no longer performed, as accurate staging is possible with modern imaging tests (helical CT and MRI).

Splenectomy: surgical aspects

The first splenectomy in the human is said to have been done by Zaccarelli of Naples in 1649, for splenomegaly in a 24 year old female. The truth of this report has been questioned, as have reports of splenectomy performed in the sixteenth and seventeenth centuries. The first successful splenectomy in the USA was reported by O'Brien in 1816 for splenic evisceration following a knife wound. In 1866, Spencer Wells performed the first elective splenectomy in the UK. Common knowledge at that time held that the spleen was expendable and splenectomy had no untoward side effects. Laparoscopic splenectomy was first reported in several European and North American centres in 1989–1990.

Preoperative management and preparation

As previously mentioned, all patients undergoing elective splenectomy should be immunized against *Streptococcus pneumoniae and Haemophilus* spp. and this should be carried out 2 weeks before surgery. Otherwise, preoperative preparation for splenectomy should be routine as for any major abdominal operation and this includes prophylaxis against thromboembolic disease. Special consideration should be given to the patient's haematological findings, clotting parameters and liver enzymes. Patients with bone marrow dysfunction or with immune platelet destruction abnormalities may be markedly thrombocytopenic prior to operation. If the platelet count is low on the basis of bone marrow failure, then platelet transfusion to a level greater than 60 000 is indicated both prior to and following operation for the first few days, in an effort to prevent bleeding episodes. If the thrombocytopenia is on the basis of immune disease, e.g. immune thrombocytopenic purpura, then preoperative platelet transfusions will be less useful, whereas human IgG will increase the platelet count. Coagulopathies due to liver disease require replacement therapy with fresh frozen plasma or cryoprecipitate as determined by factor assay.

In patients with massive splenomegaly, it may be useful to consider immediate preoperative radiological embolization of the splenic artery. Otherwise, it is usually possible to ligate the splenic artery at the superior border of the pancreas via the lesser sac at the beginning of the operation. This allows for a period of 'autotransfusion' of the sequestered blood elements during the remaining dissection. As removal of a massive spleen may be accompanied by substantial blood loss, a cell-saver system for autotransfusion should be available. Particular attention should be made during the ligature and division of the splenic hilar vessels to avoid inadvertent damage to the tail of the pancreas.

Attention to exact haemostasis is critical for the reduction of postoperative complications. The bed of the spleen, especially the raw surface of the diaphragm, should be meticulously inspected for oozing and bleeders before closure. The use of the argon spray coagulation is very effective in ensuring a dry splenic bed. Drains should not be inserted after splenectomy as they are ineffective and enhance the risk of subphrenic infection.

Laparoscopic splenectomy

Because of its advantages (fewer perioperative complications, reduced morbidity and a shorter hospital stay), laparoscopic splenectomy has now replaced open operation for most disorders requiring splenectomy or splenic surgery. There is some evidence, however, that accessory spleens are more difficult to identify laparoscopically. This may result in a recurrence of the thrombocytopenia following laparoscopic splenectomy for idiopathic thrombocytopenia, although there are no comparative studies and the results of laparoscopic splenectomy in reversing the thrombocytopenia and for acquired haemolytic anaemia have been equivalent to those obtained by open splenectomy.

Laparoscopic splenectomy has also been performed for isolated traumatic splenic injuries although the reported experience is limited. There are some concerns in relation to laparoscopic splenectomy for neoplastic disease as fragmentation/morcellation of the specimen may promote tumour spillage and implantation and also render pathological examination more difficult.

Laparoscopic approaches

A totally laparoscopic approach is used when the diseased spleen although enlarged does not extend significantly below the left costal margin and short of the umbilical plane equivalent to a longitudinal long axis on ultrasound or CT not exceeding 15 cm. Above this size and with increasing weight (>1.0 kg) splenectomy by the total laparoscopic approach can be difficult and is attended by high conversion rates. Hence until the advent of effective devices that enable hand-assisted laparoscopic surgery (HALS), the correct and appropriate surgical treatment for these patients was open splenectomy. Nowadays, however, the majority of spleens of size ranging from 20 to 40 cm (longest diameter) can be safely removed by the HALS approach, which reduces considerably the technical difficulty of the procedure and enhances safe completion without conversion. It has replaced preoperative percutaneous embolization of the splenic artery. Thus the HALS approach is strongly recommended for all adult patients with massive splenomegaly (>20 cm on ultrasound or CT). These large spleens are most often caused by congenital haemolytic anaemias, hypersplenism (primary and secondary) and myeloproliferative disorders. As the HALS technique used differs from the classical splenectomy by the total laparoscopic approach, the two operations are described separately. It should be stressed that the postoperative morbidity of laparoscopic splenectomy is significantly higher in patients undergoing the procedure for massive splenomegaly, although there is some evidence that it is reduced when performed by the HALS approach. The hand access device favoured by the author is the Omniport (Figure 26.10) but there are others such as the Hand Port, the Lap Disk and the Gel Port.

Total laparoscopic splenectomy

The 'hanging spleen' technique is most commonly used with the patient placed just short of the right lateral position (left side up) with the operating table in a slight head-up tilt (20°) facing the surgeon and the camera person on the right, with the scrub nurse on the other side of the table by the patient's back. The operation begins with detachment of the stomach from the spleen and

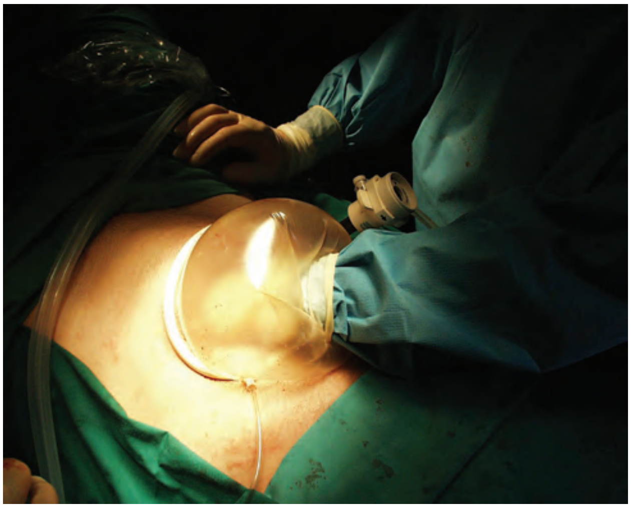


Figure 26.10 Omniport hand assist device for hand-assisted laparoscopic surgery splenectomy

division of the gastrosplenic ligament and short gastric vessels. An avascular window of the gastrosplenic ligament is identified for opening the lesser sac by scissors or ultrasonic dissection. The next step of the operation consists of dissection of the main splenic vascular pedicle which is made easier if the peritoneum overlying the pancreas is divided to expose the areolar tissue plane beneath the main splenic artery and the splenic vein. Careful dissection in this plane enables separation of the tail of the pancreas, which is at times closely applied to the main splenic vessels and risks being injured if cross stapling is carried out. Whenever possible, the author prefers to isolate the pancreatic segment of the common splenic artery along the upper border of the pancreas using curved coaxial scissors and duckbill forceps (Figure 26.11a,b); the artery is ligated in continuity intracorporeally or with an external slip knot (Figure 26.11).

The ligature in continuity of the pancreatic segment of the splenic artery achieves one important immediate benefit especially in large/massive spleens – reduction of the splenic volume aside from increase in the circulating blood volume as the splenic red pulp drains via the intact splenic vein. Additionally, it reduces the risk of major bleeding. The splenic vein is mobilized further down together with the now collapsed prehilar segment of the splenic artery (Figure 26.12a). The vein is then tied together with the collapsed prehilar artery (Figure 26.12b).

Alternatively no attempt is made to dissect and separate the artery from the vein. Instead the curved duckbill forceps or Maryland forceps is used to create the necessary space underneath the splenic pedicle lateral to the tail of the pancreas for the insertion of the linear cutting stapler mounted with vascular cartridges (Figure 26.13). To ensure accurate placement of the stapler limbs completely across the vessels, it is wise to insert a vascular sling which is then used to open the space. The stapling technique is obviously much quicker but may carry a risk of atriovenous fistula. Problems may be encountered with the stapling technique in spleens with a wide diffuse (magistral) pedicle such that more than one stapler application is necessary to secure it. Bleeding may then be encountered after the first stapler application is released. The best option to deal with this problem is suction with minimal delay before the second stapler application. The other potential problem with stapler transection of the splenic hilar vessels is risk of damage to the tail of the pancreas.

Following ligature/stapling of the vessels, the spleen is inspected closely. It should assume a uniformly dusky appearance. If not and segments are seen that are still relatively 'pink', a careful search must be made for accessory/segmental vessels, the most common of which arises from the left gastroepiploic artery (Figure 26.14).

The next step following splenic devascularization is division of the superior leaf of the lienorenal ligament. A fan or equivalent retractor is then placed on the dorsal surface of the spleen and used to displace the organ gently downwards to expose the superior peritoneal leaf of the lienorenal ligament. This is then divided with ultrasonic shears, scissors, electrosurgical hook knife or LigaSure starting at the lower pole of the spleen and including the splencolic peritoneal fold (Figure 26.15). This is followed by separation of the posterior areolar attachments of the spleen.

The fully mobilized spleen is captured in a bag such as the Endo Catch or equivalent (Figure 26.16). It usually requires finger morcellation. The neck of the bag is then exteriorized through the port wound after removal of the port. The rim of the opened bag is surrounded by Betadine swabs and the spleen fragmented by sponge forceps inside the bag and removed piecemeal.

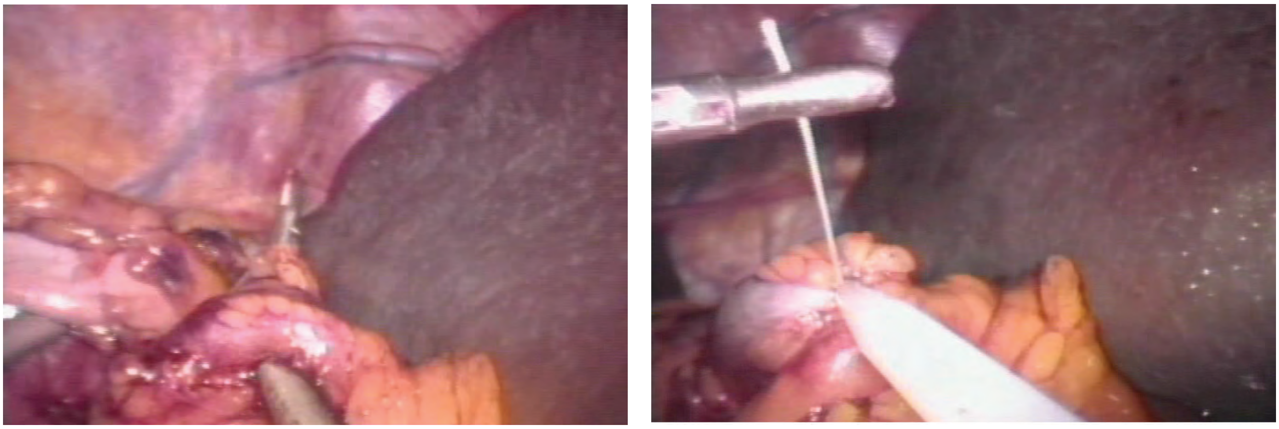


Figure 26.11 The curved coaxial duckbill forceps is under the dissected artery which is ligatured in continuity with an external slip knot. The pancreatic segment of the splenic artery may be clipped instead unless it is atherosclerotic, when clipping may be dangerous as the artery may with major bleeding requiring immediate emergency conversion.

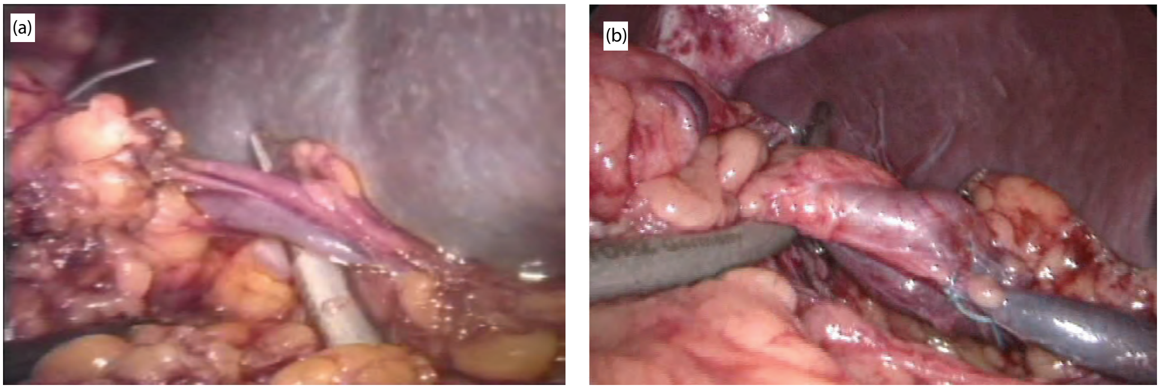


Figure 26.12 (a,b) Dissection and ligature of the splenic vein with the collapsed prehilar splenic artery.

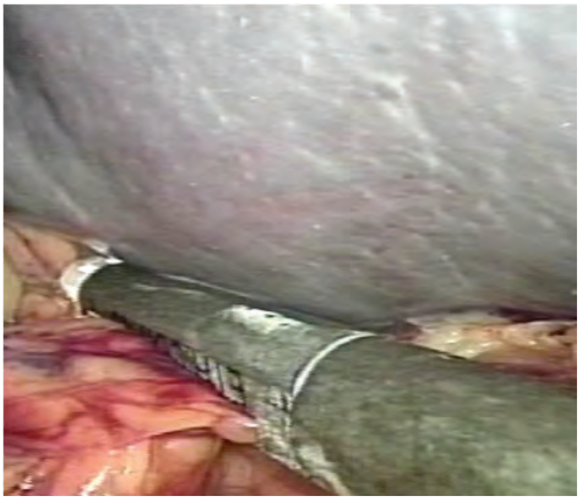


Figure 26.13 Stapling of splenic hilar vessels.

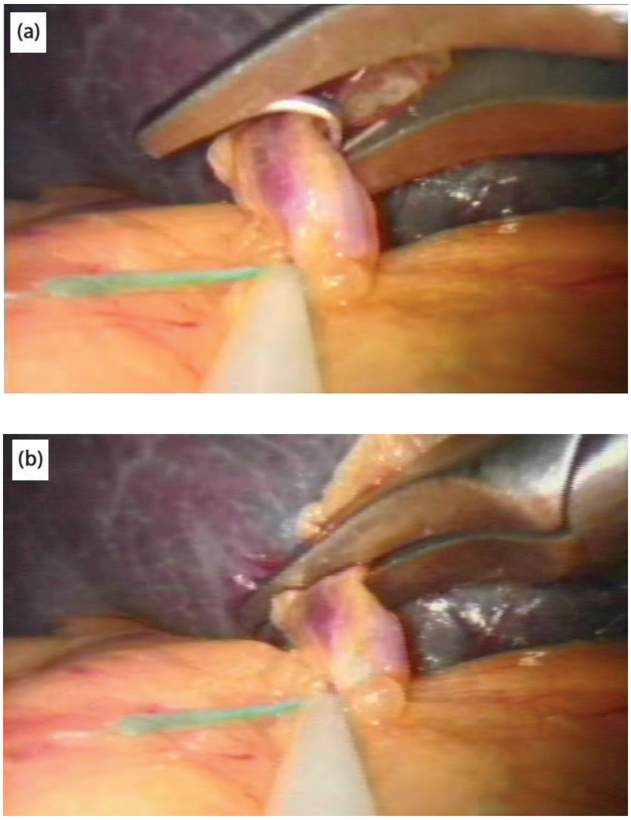


Figure 26.14 (a,b) Proximal ligation and distal clipping of a segmental polar artery arising from the left gastroepiploic artery.

HALS splenectomy

The incision for the placement of the HALS device is vertical along the linea alba at a variable distance from the umbilicus depending on the size of the spleen. The non-dominant hand is then inserted through the HALS device into the peritoneal cavity. The internal hand is used to elevate the hilar surface of the spleen to enable the insertion and deployment of a fan retractor which is then fixed to and held in position by the Martin's arm clamped to the rail along the left side of the operating table. The internal hand is then used to stretch the gastrosplenic ligament and an avascular window is identified for opening the lesser sac by scissors or ultrasonic dissection (Figure 26.17).

Once inside the lesser sac, the splenic artery is identified running sinuously along the upper border of the pancreas towards the hilum of the spleen. The stomach is then detached from the spleen by LigaSure/ultrasonic division of the gastrosplenic ligament and short gastric vessels. Whatever technique is used, the detachment of the top end of the gastrosplenic ligament can be difficult because of the enlarged splenic pole with resulting foreshortening of the ligament such that the stomach wall approximates the splenic parenchyma and can be damaged unless extreme care is taken (Figure 26.18).

The pancreatic segment of the splenic artery is exposed by division of its peritoneal and fascial coverings with the electrosurgical hook knife, and the artery then separated from the pancreas by curved duckbill or Maryland's forceps. The mobilized pancreatic segment of the splenic artery is then clipped or tied in continuity (Figure 26.19).

The division of the peritoneum overlying the pancreas and the areolar tissue attaching the spleen to the Gerota's fascia, perinephric fat and pancreas (inferior layer of the lienorenal ligament) is a crucial technical point in HALS splenectomy as it identifies the areolar tissue plane and avoids all the major branches of the splenic artery and tributaries of the splenic vein. This step is best carried out with the electrosurgical hook knife (Figure 26.20). Although largely avascular, occasional small vessels may cross this areolar plane but these are easily secured with ultrasonic dissection or clips.

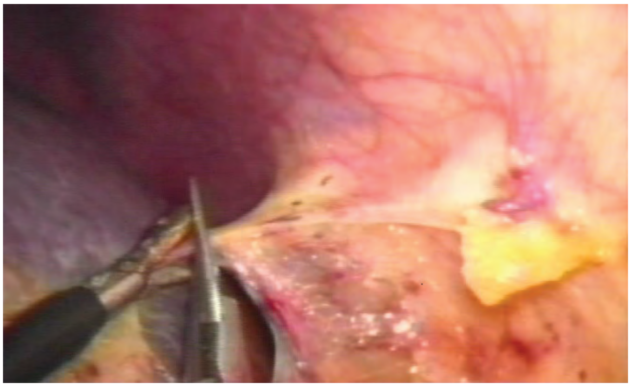


Figure 26.15 Division of the superior leaf of the lienorenal ligament starts at the lower pole and includes the suspensory ligament.

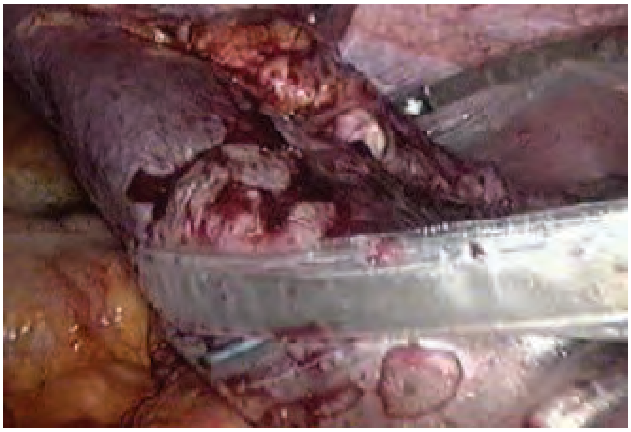


Figure 26.16 Capture of the spleen by an End Catch bag.

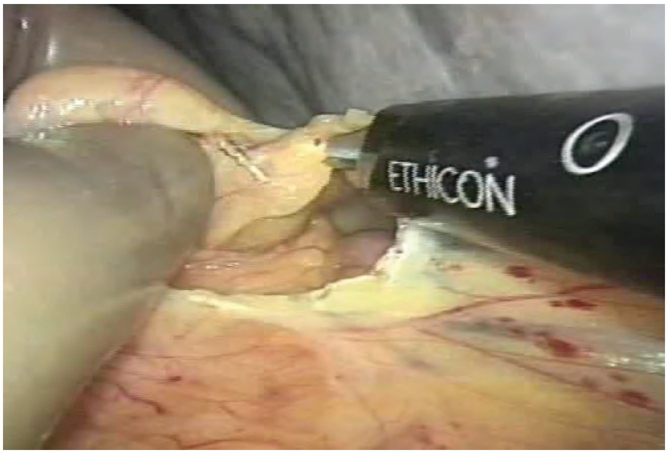
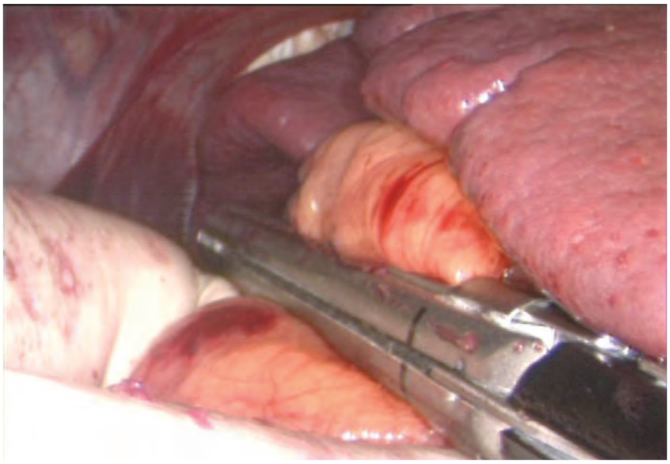


Figure 26.17 Opening the lesser sac.



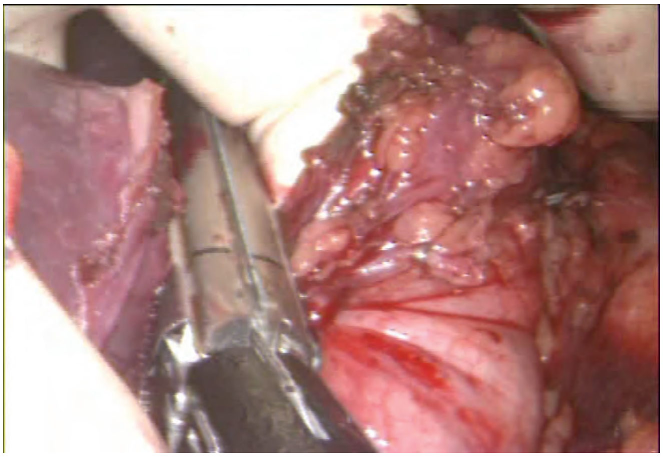


Figure 26.18 Detachment of the gastrosplenic ligament and short gastric vessels by a linear cutting angulated endostapler using a vascular cartridge. This step can be done by LigaSure (Atlas) or ultrasonic dissection.

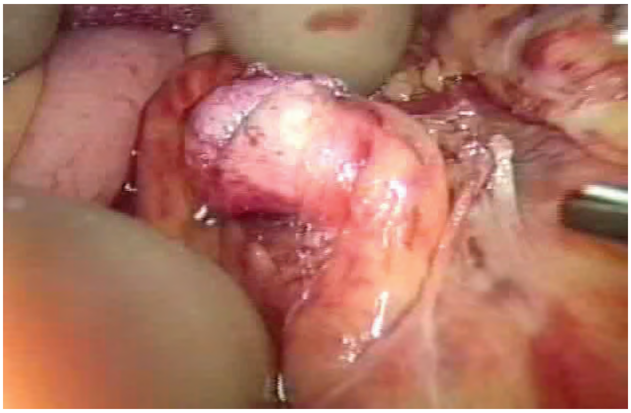


Figure 26.19 Ligature/clipping of the splenic artery in continuity.

The separation of the inferior border and surface of the spleen from the pancreas and perinephric fat is continued through the areolar tissue separating the peritoneal leaf reflected down from the hilum of the spleen (inferior leaf of the lienorenal ligament) to the perinephric fat and pancreas (Figure 26.21).

This separation must be sufficient to ensure that the pancreatic tail is not damaged when the main hilar vessels are ligated. The space created should enable the passage of a curved coaxial grasper or Maryland forceps behind the hilar vessels (arteries and veins), followed by the tip of the index finger of the dissecting hand. The splenic vessels are best tied intracorporeally. Alternatively, the vessels may be secured using external slip knots of braided 1/0 polyester.

The splenic artery and vein are then divided with curved coaxial scissors between the ligatures and the clips (Figure 26.22).

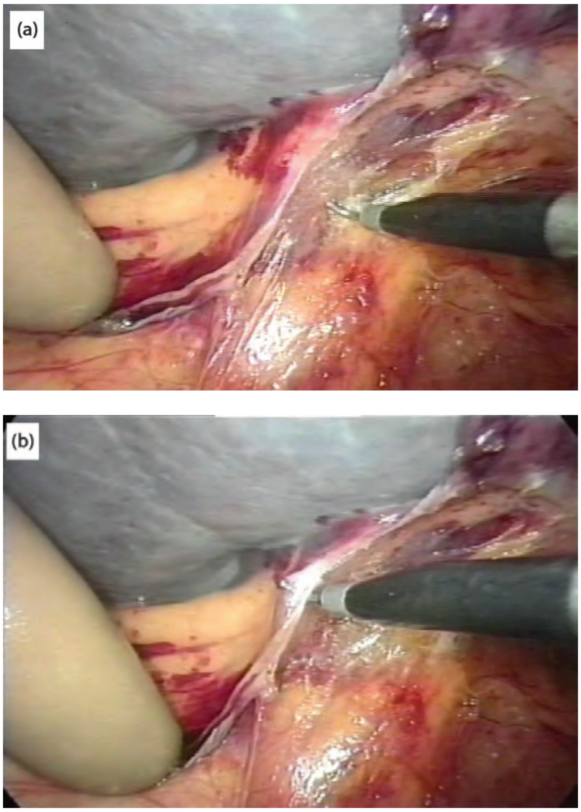


Figure 26.20 (a,b) Division of the inferior leaf of the lienorenal ligament and the areolar tissue plane binding it to the perinephric fat and pancreas.

Special measures for securing a diffuse (magistral) splenic vascular pedicle in large spleens

These can pose a major risk of bleeding, especially when the magisterial system forms a thick fat-laden bunch of interwoven arteries and veins too wide to be encompassed by the open stapler arms which cannot be made to embrace the whole pedicle. This situation requires a different technique which consists of blunt gentle dissection with the index and ring finger of the internal assisting hand in the areolar tissue plane behind the tail of the pancreas and adjoining splenic magistral vascular pedicle so that this comes to be encircled between the fingers behind and the thumb of the assisting hand in front (Figure 26.23).

Following splenic devascularization, the hand is placed on the surface of the spleen and is used to displace the 'vascularly detached' spleen gently downwards to expose the superior peritoneal leaf of the lienorenal ligament. This is then divided with ultrasonic dissection or LigaSure or an electrosurgical hook knife starting at the lower pole of the spleen and including the splenocolic peritoneal fold (Figure 26.24).

The last step (separation of the posterior areolar attachments of the spleen) can be tedious if the spleen is very large. The trick is to elevate the head of the table further and to use the fingers and thumb of the internal hand to pick up and stretch the remaining areolar tissue attachments so that they can be visualized and divided.

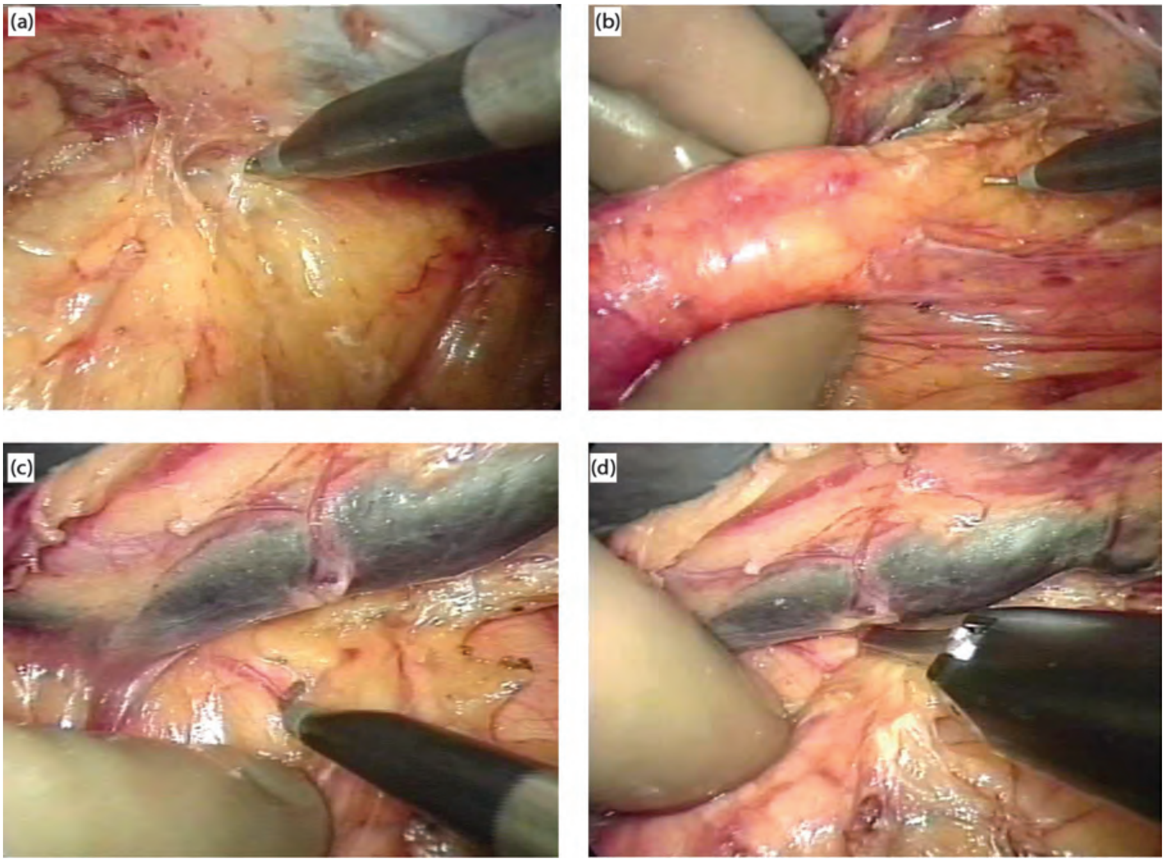


Figure 26.21 (a–d) Separation of the inferior surface of the spleen from the perinephric fat and pancreas.

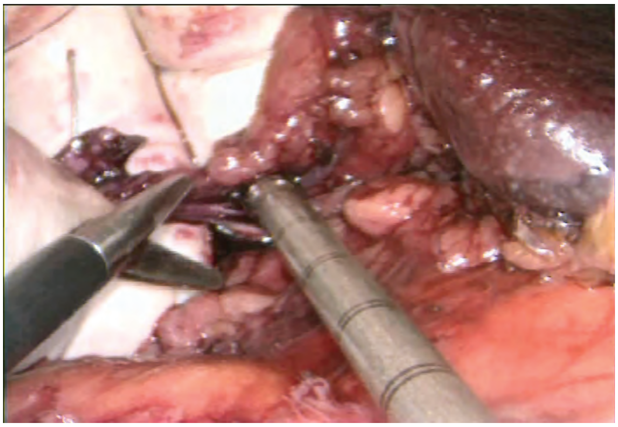


Figure 26.22 The splenic artery and vein are cut distal to the ligatures.

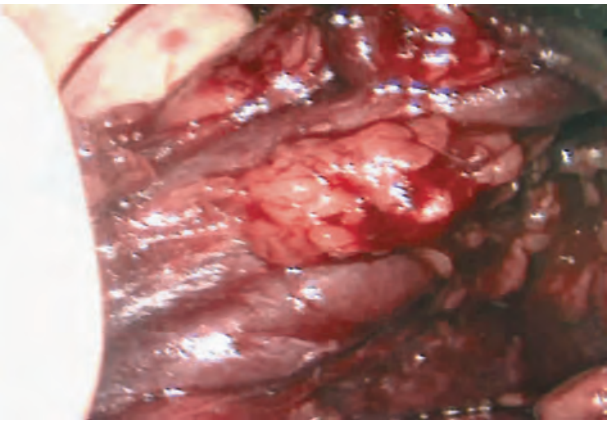


Figure 26.23 Magistral fat-laden vascular splenic pedicle with intertwining veins and arteries held and compressed for dissection between the fingers in the retropancreatic/splenic areolar plane behind and the thumb in front.

Delivery of a large spleen constitutes the most difficult part of the procedure and is also hazardous in patients in whom the splenectomy is carried out for malignancy (usually lymphoma, chronic leukaemia). Currently, the only option is to place the mobilized and completely detached spleen inside a large sterile plastic bag (e.g. used for organ donation). The problem is threefold: (1) difficulty in placing large spleens inside organ donation bags, which has to be achieved largely by the internal hand with very limited visual control; (2) during delivery of the neck of the bag through the hand access wound the large spleen often comes to lie horizontally rather than end-on, with respect to the plane of the wound; and (3) the bag is fragile and, unless great care is taken, it can be damaged during the delivery and morcellation of the splenic tissue.

The delivery steps are as follows.

1. Replacement of the HALS device with the organ retrieval bag through the wound with the base of the bag first.
2. The internalized bag is placed behind and beneath the lower pole of the spleen by the internal hand aided by atraumatic graspers introduced through one of the ports.
3. The rim of the bag is then moved proximally by the internal hand (aided by the grasper) to envelop and capture the spleen, at which point, the draw-strings are grasped by the internal and then gently exteriorized through the wound. This entire process must be unhurried and the exteriorization of the bag to approximate the spleen to the undersurface of the wound must be undertaken without any undue force. It is also a wise precaution to take out all the ports at this stage to ensure against port-induced damage to the bag during the morcellation and retrieval of the splenic tissue.
4. The area surrounding the opened neck of the bag and the wound itself is completely protected by gauze swabs soaked in povidone iodine (Betadine).
5. Morcellation is then commenced with a combination of fingers, sponge holders and scissors after a decent piece of splenic tissue is obtained for histology.
6. The morcellated splenic tissue is collected in a basin and weighed.

An alternative technique entails the use of a large plastic ring wound protector (22.5 cm). The protector is converted to a sleeve by a strong purse-string 1/0 100 mm suture (SurgiWhip, Tyco) which is tied using slip a knot. After removal of the HALS device the plastic sleeve formed from the wound protector is inserted through the wound by the push rod leaving the O ring externally and used to ensheath the spleen, at which point the purse-string is closed converting the sleeve into a closed bag holding the spleen. Approximation of the spleen inside the bag to the wound is obtained by graded withdrawal of the push rod since this lifts the closed end of the sac eccentrically up towards the parieties. Morcellation starts at this stage. As the spleen volume is reduced, tension is kept by further graded withdrawal of the push rod. Once the closed purse-string section is exteriorized, the push rod is cut to open the sac, when morcellation begins.

Complications of splenectomy

The complications of splenectomy include those common to upper abdominal operations (wound and chest infections, etc.) in addition to specific early and late complications. Major complications are encountered in 10% of patients after open splenectomy reflecting the high-risk patients with serious haematological disease undergoing this procedure. The mortality varies with the nature of the underlying disease.

Early complications

Postoperative bleeding is most commonly due to oozing from the raw surface of the diaphragm and retroperitoneum, especially after removal of large spleens and in the presence of low platelet counts or impaired platelet function. Additionally, patients with prior episodes of splenic infarction may have densely adherent yet vascular adhesions, which exacerbate postoperative bleeding. A similar problem is encountered after splenectomy in patients with portal hypertension.

Minor postoperative bleeding results in a subphrenic collection, and this complication is common in patients whose platelet count does not rise above 5×10^9/dL following the operation. Careful monitoring of the platelet counts and clotting parameters in those with hepatic dysfunction may prevent some of these collections which may become infected.

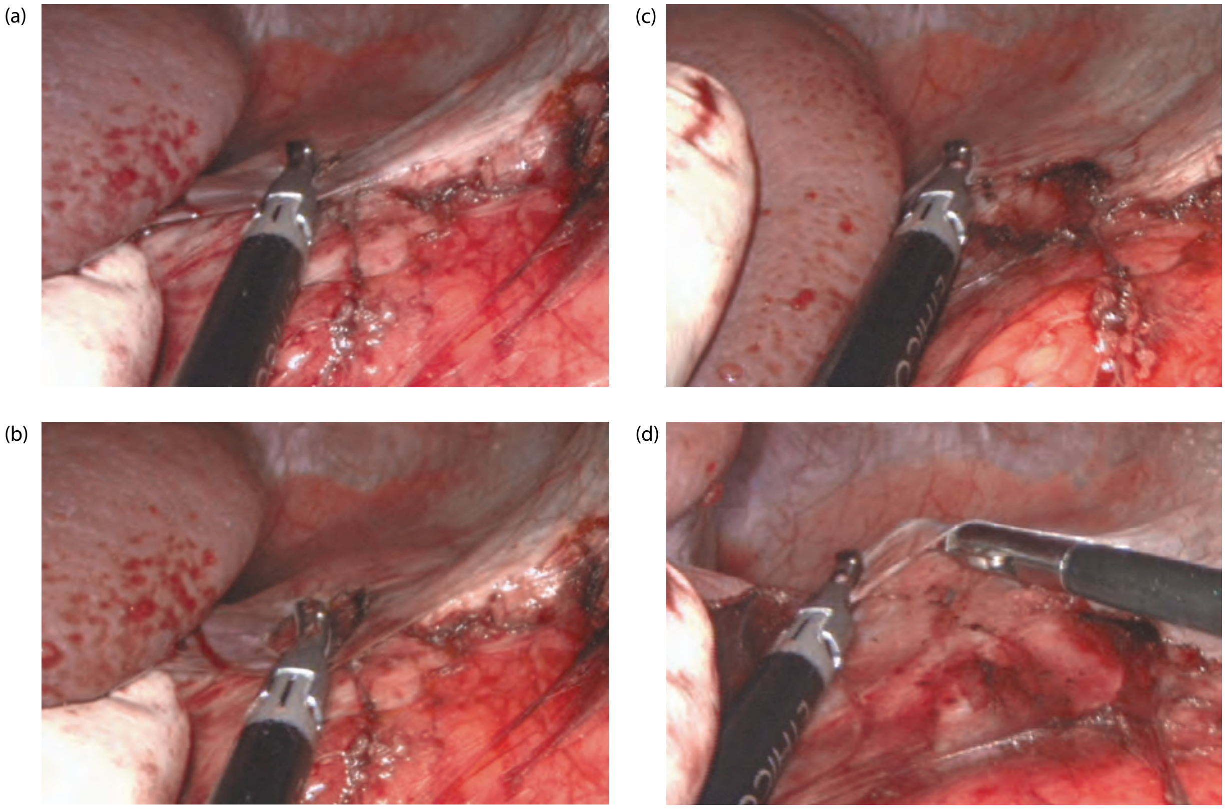


Figure 26.24 (a–d) Division of the superior leaf of the lienorenal ligament

Thrombocytosis with a hypercoagulable state may occur after splenectomy. This ‘rebound’ phenomenon is more common following splenectomy for myeloproliferative disorders. If the platelet count is in excess of 100×10^9/dL, aspirin or anticoagulants should be administered. With adequate prophylaxis (heparin and graduated compression stockings) postoperative deep vein thrombosis and pulmonary embolism should be minimized.

Iatrogenic trauma to the surrounding viscera can occur during splenectomy. During the ligation of the short gastric vessels, if the stomach wall is inadvertently included in the ligature, necrosis of the gastric wall with subphrenic abscess and gastric fistula formation will ensue. The tail of the pancreas is always intimately associated with the hilum of the spleen. For this reason, extreme caution must be exercised during dissection and ligature of the splenic hilar vessels. The current widespread practice, especially during laparoscopic splenectomy, of stapling the splenic arteries and veins together is particularly prone to this complication. Furthermore, this technique incurs the late risk of arteriovenous fistula. Surgical trauma to the tail of the pancreas during splenectomy may result in pancreatic ascites, subphrenic fluid collection/abscess and pancreatic fistula. Complications such as these can be life-threatening in an already debilitated patient with serious underlying haematological disease. Prompt recognition with percutaneous CT/ultrasound-guided drainage with appropriate antibiotic therapy and somatostatin (in the case of pancreatic ascites/fistula) is essential in these patients. A less common but equally disastrous complication is damage to the splenic flexure of the colon, which results in subphrenic abscess formation and faecal fistula. There is a documented high incidence (6%) of portal vein thrombosis after splenectomy for massive splenomegaly, and for this reason postoperative Doppler ultrasound scanning is recommended in these patients. When confirmed fully anticoagulation is instituted initially with heparin and subsequently with warfarin.

Late complications

A small number of patients develop migratory thrombophlebitis or complications of deep vein thrombosis late after splenectomy. This is more likely to occur following splenectomy for haemolytic anaemia or myeloproliferative disorders. Especially in those patients in whom the anaemia does not respond (much like an inverse relationship), thrombocytosis and associated complications may ensue. These patients may require long-term anticoagulant therapy. Late recurrence of the disease may complicate splenectomy. The usual cause of recurrent anaemia or thrombocytopenia is hypertrophy of a missed accessory spleen. If suspected, the accessory spleen can be imaged with a radiolabelled nuclear scan, and curative surgical removal. The most important late complication is, however, postsplenectomy overwhelming sepsis, which was discussed earlier.

Miscellaneous disorders of the spleen

Splenic abscess

Splenic abscess may result from contiguous spread of infection from neighbouring viscera, e.g. diverticulitis or via infection of an infarcted spleen or post-traumatic splenic haematoma. Multiple abscesses may occur in the end stage of overwhelming sepsis or with bacterial endocarditis, in patients with leukaemia and in premature infants. Multiple splenic abscesses are often fatal, even with aggressive management.

The recognition of splenic abscess is commonly delayed and for this reason carries a significant mortality. It should always be suspected in patients with established splenic infarction whose pain, fever and signs do not settle with conservative management. The clinical features of a splenic abscess are essentially non-specific, i.e. fever, pain and left upper quadrant tenderness. Splenomegaly is recognized in less than 50% of patients. The chest radiograph may show a left pleural effusion, and ultrasound scanning, an immobile hemidiaphragm and gas/debris in the abscess cavity. CT scanning is diagnostic. The management options include percutaneous drainage or splenectomy which is necessary if the abscess is large or multiple. Untreated, splenic abscesses can rupture with an invariably fatal outcome.

Treatment

When diagnosed, supportive care and parenteral broad-spectrum antibiotics active against Gram-negative bacteria (usually mixed infection) and intravenous fluid therapy are instituted. Although medical management as the only treatment has been advocated, this does not reflect the consensus view as many patients with splenic abscess often have other intra-abdominal contiguous infections. In immune-compromised patients infections with mycobacterial species, *Candida* and *Aspergillus* are well documented. There are three options used in the definitive treatment of splenic abscess: (1) percutaneous drainage, (2) open or laparoscopic splenectomy and (3) open drainage. Radiologically guided percutaneous drainage is indicated for accessible uniloculated or biloculated abscesses and for surgical patients at very high risk unfit for general anaesthesia and surgery. The procedure however carries risks of iatrogenic injury of the spleen, splenic flexure, stomach, left kidney in addition to haemorrhage, empyema and pneumothorax

Splenectomy is the standard treatment of splenic abscess and in experienced hands can be performed laparoscopically. Rarely, dense perisplenic adhesions preclude splenectomy, when open splenotomy to drain the collection is the only option. Open drainage, when selected as the appropriate treatment, can be performed by one of three approaches:

- transpleural – after resection of the twelfth rib in the posterior axillary line and drainage of the abscess through the diaphragm
- abdominal extraperitoneal – abscess drained through the lateral abdominal wall and between the peritoneum and the flat abdominal muscles
- retroperitoneal – used when the abscess extends to the flank.

Complications

The complications of untreated splenic abscess (missed or delayed) include free rupture into the peritoneal cavity with generalized peritonitis, rupture into the colon and erosion of the abscess through the diaphragm. The mortality in these patients approaches 100%.

The complications which may follow treatment include:

- life-threatening haemorrhage from the splenic parenchyma or hilar vessels
- pneumothorax
- left-sided pleural effusion
- subphrenic abscess
- perforation of the colon, stomach or small intestine
- pancreatic pseudocyst or fistula
- overwhelming, postsplenectomy sepsis
- atelectasis and pneumonia.

Ectopic (wandering, floating, pelvic) spleen

Ectopic spleen is a rare condition, which is not genetic and seems to occur more commonly in women (7:1) between 20 and 40 years of age and is also encountered in children. It is due to lax attachments of the spleen to the retroperitoneum and long splenic vessels, such that the spleen 'wanders' and may become located in the lower abdomen, most commonly in the pelvis. The diagnosis of the condition is difficult in view of its rarity. It may present with an abdominal or pelvic mass with or without recurrent attacks of pain presumably due to torsion which resolves. However, by far the commonest presentation is with sudden acute abdominal pain due to torsion (which may progress to infarction; Figure 26.25). Other patients may present with hypersplenism due to congestion associated with an abdominal or pelvic mass. The diagnosis is confirmed by imaging studies, CT and duplex ultrasonography being the preferred modalities. Treatment of ectopic spleen is operative and consists of splenopexy (non-infarcted spleen). Splenectomy is only indicated (usually as an emergency) if splenic blood flow cannot be restored after detorsion of the splenic pedicle. If the spleen is viable, then splenic fixation is indicated. A variety of techniques has been described for splenopexy of the wandering spleen, if the condition is diagnosed before infarction when splenectomy is necessary. These include:

- suturing the splenic capsule to the left upper quadrant – difficult and unreliable
- placement of the spleen in a posterolateral extraperitoneal pocket at the level of the twelfth rib – safe and effective

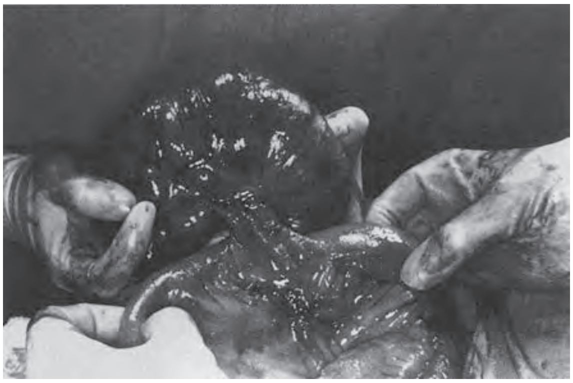


Figure 26.25 Infarcted wandering spleen. The patient presented as an emergency with acute upper abdominal pain

- positioning the left transverse colon in front of the replaced spleen, and suturing the greater curvature of the stomach to the anterior abdominal wall
- use of a polyglycolic mesh to anchor the spleen
- when the ectopic spleen is adherent to the greater omentum, this is used to anchor the organ in the left upper quadrant – possible only in cases with fortuitous adherent omentum.

Splenic cysts

Splenic cysts are primary or secondary. The latter constitute the common ones and these develop after splenic injuries, especially when treated conservatively, hence the name traumatic splenic pseudocysts (no lining epithelium).

Splenic traumatic pseudocysts may be totally asymptomatic but they have a tendency to enlarge although the natural history of these lesions is not fully understood. The time interval between the initial injury and presentation or diagnosis is extremely variable and cases are reported when this exceeded 30 years. Some traumatic pseudocysts of the spleen remain asymptomatic and are discovered accidentally during investigation by ultrasound. Others develop abdominal pain and a palpable mass or non-specific symptoms. Acute presentation with rupture is well documented. Surgical treatment is only necessary for large symptomatic cysts after confirmation of the diagnosis by ultrasound or CT. Spleen-preserving excision is possible unless the cyst is very large or presents acutely with rupture and bleeding.

True splenic cysts have an epithelial lining. There are several types:

- pseudocysts (80%) due to trauma or embolic infarction with subsequent liquefaction
- congenital epidermoid (epithelial) cysts (25%): contents may be clear or viscous from previous haemorrhage and contain proteinaceous material and cholesterol crystals
- epithelial in an accessory intrapancreatic spleen
- dermoid: rare subtype of congenital epithelial
- lymphangiomatous
- mucinous cystic lesions
- parasitic – hydatid (8%): may calcify
- necrotic splenic metastasis (rare): breast, pancreas, ovarian, endometrial, melanoma.

The commonest are pseudocysts (no epithelial lining) followed by epithelial cysts which are (Figure 26.26) lined with epithelium and are filled with straw-coloured to brownish fluid, but this may be dark and viscid following previous haemorrhage. Immunohistochemical studies have shown that the epithelial cells of the cysts express keratin, epithelial membrane antigen, carcinoembryonic antigen (CEA), carbohydrate antigen (CA) 19-9 but not BerEP4 (positive in cells of epithelial origin). These findings suggest that the origin is from epithelial metaplasia of the mesodermal undifferentiated cells from exposure to an unidentified irritant. Epithelial cysts can occur in both children and adults and can reach a large size. Some present with splenomegaly, others with left upper quadrant pain and/or non-specific symptoms (fever and non-bilious vomiting) and many are discovered accidentally during ultrasound scanning. Some have elevated serum levels of CEA and CA19-9. Small asymptomatic cysts do not require treatment but should be followed up by serial ultrasound.

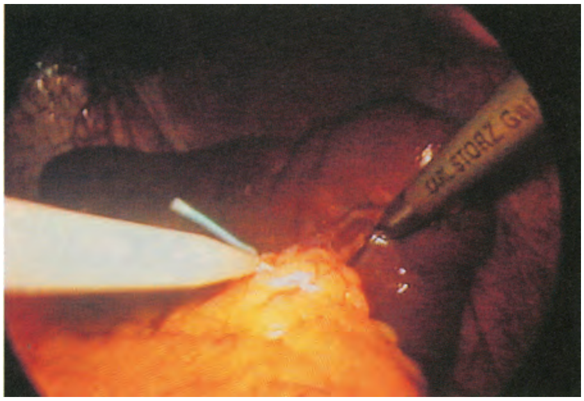


Figure 26.26 Laparoscopic splenectomy for epithelial cyst of the upper pole of the spleen.

Larger symptomatic cysts require treatment, and, unless the cyst is very large, this should consist of partial splenic decapsulation with preservation of the spleen, especially in children. Similar epidermoid cysts have been reported in intrapancreatic accessory spleens. These are also lined by non-keratinizing stratified squamous epithelium and can be multilocular. They are thought to arise from embryonic inclusion cysts of the mesothelium.

Lymphangiomas of the spleen may occur as part of lymphangiomatosis or may be solitary lesions. Solitary splenic lymphangiomas tend to form subcapsular, multicystic proliferations that are often incidental findings. Some are not true lymphangiomas and immunohistochemical studies suggest a mesothelial derivation. Splenic lymphangiomas have a distinctive CT appearance with multiple, low-attenuation lesions that do not enhance with intravenous contrast material or give a 'mottled spleen' appearance. They rarely give rise to symptoms or splenomegaly and do not require treatment if the diagnosis is certain.

All mucinous cystic lesions of the spleen are associated with malignant pseudomyxoma peritonei (peritoneal mucinous carcinomatosis) with the immunophenotype (cytokeratin 7 negative; cytokeratin 20 and CEA positive) of the cystic splenic lesions indicative of a gastrointestinal primary, most commonly the appendix (Chapter 21). The intrasplenic mucinous epithelial lesions cause splenomegaly and may thus be the presenting feature of malignant pseudomyxoma peritonei or develop as evidence of recurrent disease.

Congenital or epithelial splenic cysts

Congenital or epithelial splenic cysts are mainly seen in children and young adults. They are usually solitary, but can be multiple. Although their exact pathogenesis is unknown, several mechanisms have been proposed: involution of pluripotent cells in the splenic parenchyma during development with subsequent squamous metaplasia, entrapment of peritoneal endothelial cells or coelomic mesothelium within the developing spleen, and invagination of the surface mesothelium or dilatation of normal lymph spaces.

The cyst may be lined with columnar, cuboidal or squamous epithelium. Epithelial cysts are further subdivided as dermoid, mesothelial and epidermoid, with dermoid cysts being extremely rare. Congenital epithelial splenic cysts are usually asymptomatic. Occasionally, they cause symptoms as they enlarge often following trauma or haemorrhage from the cyst wall. Complications include infection, rupture and haemorrhage. The treatment of symptomatic splenic cyst is partial or total splenectomy.

On ultrasound scanning splenic epithelial cysts appear as round, homogeneous, anechoic lesions with a smooth thin wall, although septation, irregular cyst walls, internal debris or haemorrhage and calcifications may contribute to an inhomogeneous appearance. On CT, splenic cysts are typically spherical well-defined lesions with an attenuation value close to water and a thin or imperceptible wall with no enhancement after an injection of contrast. On MRI, the cyst is hypointense on T_1 weighted images and strongly hyperintense on T_2 weighted images, with a signal intensity equal to that of water and no enhancement after injection of contrast.

Splenic tumours

Apart from lymphomas, both primary and secondary tumours of the spleen are rare. The vascular tumours include primary angiosarcoma, haemangioma, haemangioendotheliomas and benign vascular neoplasms with myoid and angioendotheliomatous features. Splenic angiosarcomas constitute less than 1% of all sarcomas and less than 100 cases have been reported. They can present acutely with severe abdominal pain and intraperitoneal bleeding from spontaneous rupture (30%) and carry a uniformly poor prognosis.

Secondary tumour deposits are equally rare with a reported frequency of 2–5%. Although all malignant tumours can metastasize to the spleen, the most frequent site of the primary is the breast, lung, pancreas and ovary. Cutaneous melanoma has also been documented to metastasize to the spleen, and direct involvement from pancreatic and retroperitoneal sarcomas can occur.

Benign tumours of the spleen

Most common primary tumours of the spleen are benign and originate from the vascular endothelium.

Haemangiomas

Haemangiomas are the most common benign tumour of the spleen and are commonly found incidentally as the majority are asymptomatic. Most are small (less than 2 cm in diameter), but large haemangiomas carry a risk of spontaneous rupture with haemorrhage and therefore require treatment even if asymptomatic. Splenic haemangiomas are usually solitary but may be multiple or associated with haemangiomas of other sites (angiomatosis). Two principal types are recognized: the cavernous variety (most frequent) and capillary haemangioma.

Capillary haemangioma is hyperechoic on ultrasound scanning, whereas cavernous haemangioma appears as a heterogeneous hypoechoic mass, sometimes containing areas of calcification or multiple cystic spaces. On CT, capillary haemangiomas appear as small homogeneous iso- or hypodense masses with homogeneous contrast enhancement. Cavernous haemangiomas are usually bigger and cystic with occasional iso- or hypodense areas following injection of contrast. Calcifications when present may be either peripheral and curvilinear or central.

Hamartoma

Hamartoma is a rare benign tumour of the spleen, with a reported autopsy incidence of 0.13%. Some are diagnosed in children but the majority present in adult life. They are usually small (<3 cm in diameter), but can reach up to 18 cm in size. Histologically, hamartomas are composed exclusively of red pulp components, but may also contain cystic or necrotic components and areas of calcification. Spontaneous rupture of a hamartoma with acute abdominal pain in adults has been reported, but most patients are asymptomatic and the lesion is discovered incidentally during routine investigations. Ultrasound scanning may show a solid mass, which is usually heterogeneous with multiple hyperechoic areas due to punctate calcifications and cystic changes. On CT, hamartomas appear as well-demarcated, solid, hypodense masses. MRI demonstrates a well-defined homogeneous mass which is isointense on T_1 weighted images.

Littoral cell angioma

Littoral cell angioma (LCA) is a rare benign tumour of the spleen which originates from the lining cells of the red pulp sinuses (littoral cells). Although considered benign, there are well-documented reported cases with disseminated disease but these are nowadays regarded as littoral cell haemangioendotheliomas. The majority of patients are middle aged and present with anaemia, pyrexia of unknown origin and a variable degree of splenomegaly causing hypersplenism. However, many cases are asymptomatic and the exact diagnosis is only established after splenectomy. LCAs are invariably located within the red pulp of the spleen and are usually multinodular. Histologically, they are composed of anastomosing vascular channels with irregular lumen featuring cyst-like spaces and lined by tall endothelial cells.

Ultrasound scanning depicts a lobulated splenomegaly with heterogeneous echogenicity.

Lymphangioma

Lymphangioma is relatively uncommon and may present as a solitary splenic mass or as part of systemic lymphangiomatosis usually in young patients. The majority are cystic but solid lymphangiomas with sclerotic change and papillary endothelial proliferation have been reported. Cystic lymphangioma is characterized by a honeycomb of large and small thin-walled cysts containing lymph-like clear fluid. Ultrasound scanning shows multiple hypoechoic cysts of various sizes with hyperechoic septa, debris and areas of calcification. Colour Doppler ultrasound shows a fan-shaped displacement of the intrasplenic vessels extending from hilus to periphery.

Inflammatory myofibroblastic tumour (pseudotumour, plasma cell granuloma)

Inflammatory myofibroblastic tumour (IMT) is a fibroinflammatory process mainly occurring in liver and spleen. It is characterized histologically by spindle cell proliferation mixed with inflammatory cells consisting of lymphocytes and plasma cells. The proliferating spindle cells show a myofibroblastic phenotype, hence the current favoured designation as IMT. However the condition encompasses several entities with different aetiologies. Thus some IMTs are reparative, others are associated with infection by the Epstein–Barr virus, and some are neoplastic and can recur after excision and even metastasize. At presentation, the tumour may vary in size from 3.5 to 20 cm, with a median of 10 cm.

On ultrasound, IMTs appear as well-circumscribed hypoechoic lesions. CT shows a hypodense mass with delayed enhancement and a central scar. MRI shows hypointense lesions on T_1 weighted images and hyperintense lesions on T_2 weighted images.

Splenic vein thrombosis

This under-recognized condition occurs most often following either acute or chronic inflammation of the pancreas. It is also seen in infants who have had umbilical vein catheterization or from direct extension of a pancreatic neoplasm.

Isolated splenic vein thrombosis (i.e. without portal vein thrombosis) results in splenomegaly and sectorial or left-sided (sinistral) portal venous hypertension. This condition is characterized by the development of varices involving the short gastric and gastroepiploic veins. The most common presentation is with bleeding gastric varices in patients with normal or good liver function (Figure 26.27). As oesophageal varices may be absent, central portal pressures normal and gastric varices are easily overlooked at endoscopy, this condition can be missed. For this reason, gastrointestinal bleeding in the setting of prior pancreatitis should always raise the suspicion of splenic vein thrombosis. The diagnosis can be confirmed by high-dose helical CT or duplex ultrasonography, and these imaging tests have replaced direct visceral angiography and splenoportography.

Recognition of splenic vein thrombosis is critical, as otherwise the gastrointestinal haemorrhage may prove fatal. The treatment is splenectomy. Venous shunt procedures are contraindicated as central portal hypertension is not a feature of the disease, and the sectorial hypertension is not relieved by any type of central portosystemic shunt.

Splenosis

The peritoneum of patients who have sustained rupture or trauma to the spleen can be 'seeded' with splenic fragments that autotransplant. This pathological curiosity led to attempts at salvaging splenic function by intentional transplantation of fragments into the omentum. That such transplants survive is unchallenged, as demonstrated by animal studies, but the surveillance function of the spleen is lost. As such, the enthusiasm for surgical autotransplantation has declined in recent years.

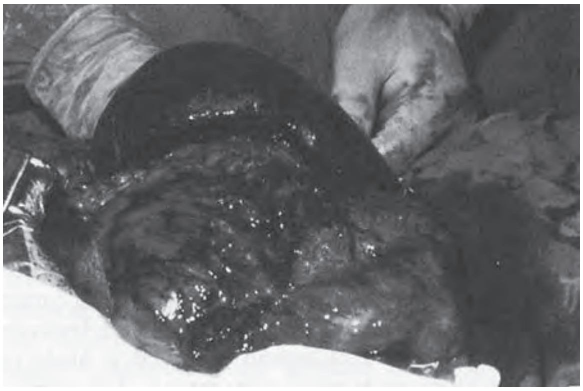


Figure 26.27 Sectorial portal hypertension caused by chronic pancreatitis with splenic vein thrombosis. The patient presented with severe haematemesis. The spleen was enlarged. Endoscopy showed bleeding gastric varices. Note the distended veins along the greater curvature and fundal region of the stomach (short gastric and left gastropepiploic).

Differentiation between splenosis and accessory spleens is not usually difficult. Accessory spleens are found in recognized locations, most often at the hilum of the spleen, tail of pancreas and omentum. They number less than 10 and have hilar vessels with normal splenic architecture. The splenic implants of splenosis often number more than 20, have antecedent traumatic events, are scattered over the surface of the peritoneum and do not have a co-ordinated circulation. Splenosis does not require treatment.

Gaucher disease

Hypersplenism and massive splenomegaly can occur due to this hereditary disorder of lipid metabolism. The three types of the disorder are due to a specific inherited enzyme defect (β-glucocerebrosidase). Type I is the most commonly encountered adult form, type II is the routinely fatal infantile form and type III the intermediate form. Patients surviving into adulthood exhibit splenomegaly, thrombocytopenia, brownish discoloration of the limbs and pingueculae. The typical Gaucher's fat-laden cells can be found on marrow aspirate, and long-bone abnormalities occur in approximately 50% of patients. Splenectomy is indicated for symptoms or complications of hypersplenism.

Disorders of lymph nodes

Palpable subcutaneous lymph nodes should be considered diseased until proved otherwise. The enlargement (lymphadenopathy) may be a transitory self-limiting process, or the harbinger of a life-threatening illness. Lymph node enlargement may be localized (regional) or generalized. It may be purely reactive (enlarged lymph nodes exhibiting sinus histiocytosis) or the nodes may be diseased (infection or involvement by tumour).

Localized lymphadenopathy

Many acute bacterial and viral infections produce localized tender lymphadenopathy. With streptococcal or staphylococcal infections the overlying skin and subcutaneous tissue is often inflamed and oedematous, although antibiotic therapy usually prevents progression to suppurative lymphadenitis. By contrast, the periadenitis is minimal in lymphadenitis caused by viral disease, although the enlarged nodes are usually tender, e.g. infectious mononucleosis.

Chronic infections and parasitic infestations may produce considerable lymphadenopathy without or with minimal signs of acute inflammation. Primary bovine tuberculosis may produce a chronically enlarged group of matted deep cervical lymph nodes that suppurate and form a 'cold' collar stud abscess as it penetrates the investing layer of the deep cervical fascia, and eventually discharges through the skin forming a tuberculous sinus. Syphilis, leprosy, fungal infections and lymphogranuloma venereum can all produce chronic indolent lymphadenopathy.

The site of the enlarged lymph nodes may be of diagnostic help. Enlarged occipital nodes usually indicate a chronic scalp infection and posterior auricular node enlargement is common in rubella. Anterior auricular lymphadenopathy is most often bacterial in origin from infection of the eyelids and conjunctivae. Cervical painless lymphadenopathy is a frequent presentation of nasopharyngeal cancer and the commonest presentation of both Hodgkin's and NHLs. Axillary adenopathy is common in breast cancer and in upper limb infections, and less frequently Hodgkin disease. Painless epitrochlear lymphadenopathy is commonly seen in childhood viral illness, secondary syphilis or generalized tuberculosis, but is rare in sarcoidosis. Palpable small 'shotty' lymph nodes in the groin do not necessarily signify disease, as they are often prominent in children and thin athletic adults. Larger fleshier nodes however should raise suspicion of disease and merit investigation.

In the chest, mediastinal or hilar lymph nodes do not become noticeably enlarged on the chest radiograph with bacterial and viral pneumonias, but pulmonary tuberculosis can produce unilateral hilar lymphadenopathy. Infectious mononucleosis may cause persistent mediastinal lymphadenopathy lasting for several months, but the most common causes of hilar lymphadenopathy are bronchial carcinoma and sarcoidosis.

Regional intra-abdominal lymphadenopathy is of paramount importance in all intra-abdominal cancers and may be due to sinus histiocytosis or secondary deposits (lymph node-positive cancers). Secondary involvement of the lymph nodes (on pathological staging) is an important feature of the pathological staging of all cancers. It affects the prognosis and may indicate the need for adjuvant therapy.

Generalized lymphadenopathy

When there is noticeable lymph node enlargement in more than one drainage region, the most common cause is a viral infection. Common viral illnesses causing generalized lymphadenopathy include infectious mononucleosis, viral hepatitis, influenza, cytomegalovirus infection, rubella, infectious lymphocytosis and AIDS. Excision biopsy of the affected node or nodes is often useful in AIDS patients in directing supportive therapy. The findings in AIDS patients having lymph node biopsy, in decreasing order of frequency, include tuberculosis, Kaposi syndrome, reactive hyperplasia, *Cryptococcus*, *Mycobacterium avium* complex, lymphoma and lymphoepithelial cysts.

Fever and generalized lymphadenopathy occurs in patients with secondary syphilis, acute leptospirosis, salmonellosis, typhoid, paratyphoid and generalized haematogenous tuberculosis. Protozoal infections, e.g. toxoplasmosis, can resemble illnesses such as infectious mononucleosis.

Non-infective and non-neoplastic illnesses producing generalized lymphadenopathy include the autoimmune haemolytic anaemia, collagen vascular disorders, hypersensitivity reactions, hyperthyroidism and a variety of skin disorders, e.g. exfoliative dermatitis.

Lymph node imaging studies

The early attempts at lymph node imaging to detect nodal involvement were performed in relation to staging of Hodgkin disease and NHLs using pedal lymphangiography. The lymphatics in the webs of the toes were identified after intradermal injection of a vital blue dye and then cannulated for contrast injection. In this fashion, proximal pelvic and abdominal lymph nodes could be imaged and involvement detected (Figure 26.28).

This method of staging of lymphomas is rarely used nowadays, but the related technique of *sentinel lymph node mapping* is practised and has been shown to provide accurate nodal staging in patients with melanoma, and, more recently, in patients with breast cancer. The original technique described by Morton *et al.* used a vital blue dye injected intradermally at the periphery of the primary tumour and this resulted in the sentinel node becoming stained blue and thus easily identified with minimal dissection. The technique has been modified by use of ^{99m}Tc-labelled human albumin colloid, 0.3 mL of which is injected intradermally at the edge of the tumour. Lymphoscintigraphy is then performed using an external dual-head gamma camera at 30 minutes and 2 hours and the site of the sentinel node marked on the skin. A special sterile radioprobe connected to a Geiger counter is also used to aid in precise location of the node during

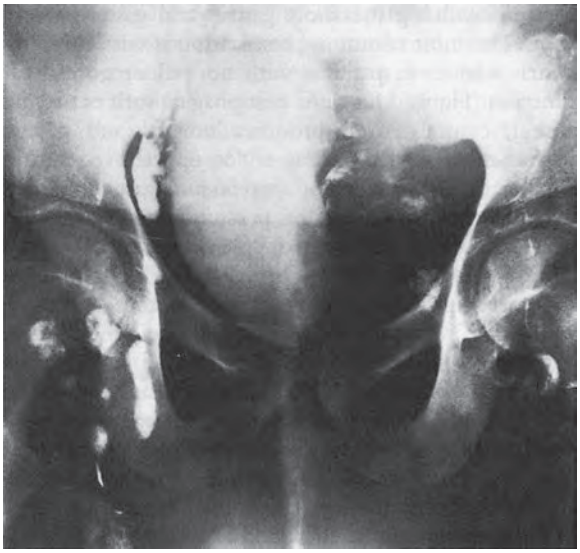


Figure 26.28 Lymphangiogram and intravenous pyelogram in a patient with non-Hodgkin's lymphoma. The very large lymph nodes on the left have filled poorly and are demonstrated as large masses indenting the urinary bladder.

surgical harvest. The combined technique where the vital blue dye is mixed with the radiolabelled albumin colloid gives the best results. Sentinel node biopsy has been shown to yield an accurate assessment of the nodal status of the regional lymph nodes and to provide a strong prognostic factor. It carries a false-negative rate of around 10% , is well established in the staging of breast cancer and is used in some centres for other tumours, e.g. melanoma, stomach, colon.

Hodgkin disease: clinical features

Hodgkin disease typically presents in middle-aged male adults. The neck is the most common site of primary presentation, as a group of painlessly enlarged anterior cervical nodes. Because of the dominant lymphatic drainage, the left side is the more common side of presentation. Axillary nodes are the site of presentation in less than 20% of patients, with the mediastinal or inguinal nodes being the first site in less than 15% of patients.

Hodgkin disease is considered to be unicentric in origin and spreads via contiguous lymphatic channels to adjacent lymph nodes or lymphoreticular organs. The rate of growth of the involved lymph nodes varies between patients, and pain is uncommon in those patients with a slow-growing lymphoma. Enlarged nodes may fluctuate in size with inflammation or necrosis, but rarely shrink enough to escape careful palpation. Systemic symptoms may appear early or late, and variably include general malaise, listlessness, anorexia, weight loss, sweating, intermittent fever and pruritus. Clinically apparent hepatosplenomegaly appears late. Unchecked, the disease may involve any organ system. It is not uncommon to discover obscure involvement of the retroperitoneal, para-aortic, iliac and deep inguinal nodes, as well as diffuse infiltration of the bone marrow. In the late stages of the disease, lymphomas may spread to the meninges, pleura, thyroid, breasts, kidneys, urinary tract and gonads. Local node masses or organ involvement may produce mediastinal obstruction, neurological syndromes, e.g. Horner's, intestinal obstruction or renal failure.

The diagnosis of Hodgkin's, as with any lymphoma, begins with biopsy of a diseased lymph node. The biopsy should be excisional since the histological pattern of the disease is critical to the management of the individual patient. The corollary to exact pathological diagnosis is accurate staging of the disease, i.e. deciding on objective evidence based on staging tests which nodal areas require treatment. For all patients this includes bone marrow biopsy, full blood count and biochemistry and high-quality imaging with helical CT or MRI of the chest and abdomen. In the past, laparotomy was an integral part of staging and included lymph node harvest, liver biopsies and splenectomy. Although it yielded valuable information that altered the stage of the disease, and hence influenced the nature of the treatment, it was attended by a significant morbidity. The very significant improvement in CT and MRI has now virtually abolished the need for staging laparotomy and lymphangiography.

The clinical stages of Hodgkin disease are as follows:

- Stage I
  - Involvement of a single lymph node region, or a single extralymphoid organ or site (IE)
- Stage II
  - Involvement of two or more lymph node groups on the same side of the diaphragm, or localized involvement of an extralymphoid organ or site and one or more lymph node groups on the same side of the diaphragm (IIIE)
- Stage III
  - Involvement of lymph node groups on both sides of the diaphragm, which may be accompanied by localized involvement of an extralymphoid organ (IIIE), spleen (IIIS) or both (IIISE)
- Stage IV
  - Diffuse involvement of one or more extralymphoid organs with or without associated lymph node involvement: sites denoted: N (lymph nodes), H (liver), M (marrow), P (pleura), S (spleen), L (lung), O (bone), D (skin)

A = asymptomatic
B = >10% body weight in 6 months, unexplained fever above 38°C, or night sweats.

Stages I–IV are further subdivided into asymptomatic patients (A), and those with systemic symptoms (B). Clinically apparent hepatomegaly does not necessarily represent infradiaphragmatic disease unless accompanied by elevated liver enzymes or abnormal scan, but a palpable spleen nearly always represents involvement. Questionable involvement of the spleen is confirmed by CT/MRI. In the past patients deemed to be in clinical stage I or IIa were subjected routinely to staging laparotomy with splenectomy which revealed stage III disease in up to 20% of patients. As mentioned previously, this is rarely undertaken nowadays because of the increased reliability of clinical staging. Splenectomy is, however, carried out for splenic involvement as this removes a radiotherapy portal that is potentially very morbid, avoiding radiation injury to the lung, kidney and intestines in the left upper quadrant.

Non-Hodgkin's lymphomas: clinical features

NHLs are a heterogeneous group of tumours of the immune system with different origins, different natural histories and different prognoses. This diversity requires multiple treatment strategies. In contrast to Hodgkin disease, NHLs tend to present more insidiously, with up to 25% of patients harbouring abdominal complaints because of retroperitoneal adenopathy. The majority of these patients present with painless enlargement of one or more superficial node groups. Less commonly, extranodal regions are the primary site of the disease, including the skin, orbit, pituitary, thyroid, tracheobronchial tree, gastrointestinal tract and central and peripheral nervous systems. Approximately one-half of patients at presentation have no systemic complaints, but with disease progression, similar complications and symptoms may occur as in Hodgkin disease.

In 1956, Rappaport proposed a classification scheme that proved both reproducible to pathologists and prognostically relevant. This system was used through the 1970s, when several new schemes were introduced with no scheme being clearly superior. Increasingly, the joint American–European classification is used for classification of the tumours based on morphological, immunological and molecular characteristics of the tumours. Two broad types are recognized: (1) -cell lymphomas and (2) T-cell and putative NK-cell tumours. Within each category, some subtypes are still regarded as a 'provisional entity'. Unlike Hodgkin disease, the stage of the disease and spectrum of NHLs are accurately predicted on clinical grounds, and staging laparotomy has never been proven to be of value, and for this reason is not practised. Patients whose tumours show histiocytic morphology tend to develop systemic symptoms earlier than other varieties of lymphoma. Malaise, weight loss, superficial lymphadenopathy and hepatosplenomegaly occur early in this form of the disease. In 'fulminant' cases, fevers, rapid tissue wasting, massive lymphadenopathy and marked hepatosplenomegaly occur with alarming speed (Figure 26.29). There is a significant danger of spontaneous rupture of the spleen in some of the more clinically aggressive forms of NHL, e.g. histiocytic medullary reticulosis. Mycosis fungoides and Sézary syndrome are two forms of NHL that present with early cutaneous manifestations and later with generalized lymphadenopathy and/or systemic complaints.

Treatment strategies for non-Hodgkin's lymphomas and Hodgkin disease

The treatment options for Hodgkin disease and NHLs are multiple and varied. Indeed, treatment of these disorders is one of the relatively few areas where recent advances have had a major impact upon survival and cure rates (up to 90% 5 year survival for Hodgkin disease). Critical to successful management is the accurate staging of the disease. This is especially true for patients with Hodgkin disease.

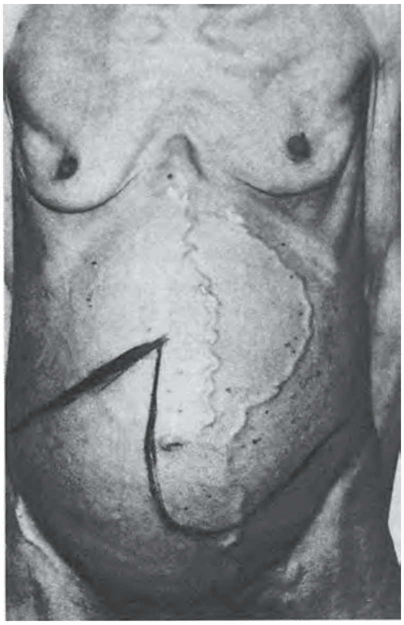


Figure 26.29 Patient with a low-grade malignant non-Hodgkin's lymphoma. Note enlargement of the left groin nodes and hepatosplenomegaly. Clinically there was obstruction of the left common iliac vein with dilatation of veins on the anterior abdominal wall.

Patients with stages I and II Hodgkin disease are treated initially with wide-field megavoltage radiotherapy. This is in the form of 'mantle' and 'inverted Y' fields to cover the known areas of the disease. Patients with IIIA and more advanced disease are treated with combination chemotherapy. The early stages of NHL are treated in similar ways to Hodgkin disease. Local radiotherapy is the first-line treatment of stage IA, or stage II with extranodal disease, bone lesions and/or some of the lymphomas of the gastrointestinal tract. In addition, NHLs of the mediastinum in children and adolescents are offered radiotherapy as a first-line therapy. Wide-field radiation is given, totalling 3000–4000 cGy over a 4 week period. Relapses occur most commonly within the first 2 years following such therapy. More advanced stages and patients with recurrent disease are given combination chemotherapy. There is no one 'standard' regimen for the treatment of NHLs, but combinations of cyclophosphamide, vincristine and prednisone are usually employed in pulses over several months. The nodular and diffuse lymphocytic varieties carry a more favourable prognosis than the histiocytic malignancies. Patients with stage I histiocytic lymphomas, therefore, usually receive primary radiotherapy to the affected region but, because of the poorer response rates, patients with stage II or greater disease are also given chemotherapy in the first instance.

All regimens of chemotherapy produce varying degrees of neutropenia and/or thrombocytopenia. Severe, life-threatening infections are frequent. All neutropenic patients need regular oral inspection and maintenance hygiene with a prophylactic fungicide mouthwash. Rectal infections must be recognized early and treated with broad-spectrum antibiotics and surgical drainage as required. Bleeding during thrombocytopenic episodes must be treated with platelet transfusions. Often, anaemia complicates the clinical picture because of bleeding complications, underlying disease and/or bone marrow suppressions. Replacement component therapy is therefore given as needed.

It should be remembered that some chemotherapeutic agents, notably vincristine, cause peripheral and autonomic neuropathies. As such, paralytic ileus with constipation is common, and should not be confused with mechanical obstruction, nor should it be ignored. Typhlitis is a condition of these patients where transmural necrosis of the caecum occurs, probably as a result of bacterial overgrowth within the setting of a paralytic ileus. Patients requiring surgery for this complication have a mortality exceeding 50%.

Mediastinal obstruction

Mediastinal lymph node involvement with lymphoma can occasionally cause mediastinal obstruction. This is particularly true of Hodgkin disease, diffuse histiocytic tumours, or undifferentiated lymphomas of childhood. Highly differentiated mediastinal tumours in children (Sternberg's sarcoma) frequently develop into acute lymphoblastic leukaemia after initial treatment.

Patients with mediastinal obstruction present a therapeutic challenge. It is very important to obtain sufficient tissue for accurate histological diagnosis (anterior mediastinotomy, mediastinoscopy, thoracoscopy). If respiratory distress complicates the presentation, urgent treatment with corticosteroids, radiotherapy or vinca alkaloids may be necessary during the staging process. Hodgkin disease presenting in this way is often associated with a good outcome. However, childhood lymphomas that are complicated by the development of acute lymphoblastic leukaemia carry a particularly poor prognosis.

GUIDE TO FURTHER READING

Al-Salem AH, Naserullah Z, Qaisaruddin S, *et al.* Splenic complications of the sickling syndromes and the role of splenectomy. *J Pediatr Hematol Oncol* 1999;**21**:401–406.

Andrews MW. Ultrasound of the spleen. *World J Surg* 2000;**24**:183–187.

Arber DA, Strickler JG, Weiss LM. Splenic mesothelial cysts mimicking lymphagiomas. *Am J Surg Pathol* 1997;**21**:334–338.

Brigden ML, Pattullo AL. Prevention and management of overwhelming postsplenectomy infection – an update. *Crit Care Med* 1999;**27**:836–842.

Cavenagh JD, Joseph AE, Dilly S, Bevan DH. Splenic sepsis in sickle cell disease. *Br J Haematol* 1994;**86:**187–189.

Desai DC, Hebra A, Davidoff AM, Schnaufer L. Wandering spleen: a challenging diagnosis. *South Med J* 1997;**90**:439–443.

Du Plessis DG, Louw JA, Wranz BPA. Mucinous epithelial cysts of the spleen associated with pseudomyxoma peritonei. *Histopathology* 1999;**35**:551–557.

Gielchinsky Y, Elstein, D, Hadas-Halpern I, *et al.* Is there a correlation between degree of splenomegaly, symptoms and hypersplenism? A study of 218 patients with Gaucher disease. *Br J Haematol* 1999;**106**:812–816.

Grotto HZ, Costa FF. Hyposplenism in AIDS. *AIDS* 1991;**5**:1538–1540.

Han, MJ, Zhao HG, Ren K, *et al.* Partial splenic embolization for hypersplenism concomitant with or after arterial embolization of hepatocellular carcinoma in 30 patients. *Cardiovasc Intervent Radiol* 1997;**20**:125–127.

Heniford BT, Park A, Walsh RM, *et al.* Laparoscopic splenectomy in patients with normal-sized spleens vs splenomegaly: does size matter? *Am Surg* 2001;**67**:854–858.

Hon TY, Chan CC, Loke T, Chan CS. Torsion of the wandering spleen. *Australas Radiol* 1998;**42**:258–261.

Jansen L, Nieweg OE, Peterse JL, *et al.* Reliability of sentinel lymph node biopsy for staging melanoma. *Br J Surg* 2000;**87**:484–489.

Kercher KW, Matthews BD, Walsh RM, *et al.* Laparoscopic splenectomy for massive splenomegaly. *Am J Surg* 2002;**183**:192–219.

Khan AM, Harrington RD, Nadel M, Greenberg BR. Hyposplenism from Mycobacterium avium complex infection in a patient with AIDS and immune thrombocytopenia. *Acta Haematol* 1998;**99**:45–48.

Klingler PJ, Tsiotos GG, Glaser KS, Hinder RA. Laparoscopic splenectomy: evolution and current status. *Surg Laparosc Endosc* 1999;**9**:1–8.

Nores M, Phillips EH, Morgenstern L, Hiatt JR. The clinical spectrum of splenic infarction. *Am Surg* 1998;**64**:182–188.

Powsner RA, Simms RW, Chudnovsky A, *et al.* Scintigraphic functional hyposplenism in amyloidosis. *J Nucl Med* 1998;**39**:221–3.

Pursnani KG, Sillin LF, Kaplan DS. Effect of transjugular intrahepatic portosystemic shunt on secondary hypersplenism. *Am J Surg* 1997;**173**:169–173

Sasou S, Nakamura SI, Inomata M. Epithelial splenic cysts in an intrapancreatic accessory spleen and spleen. *Pathol Int* 1999;**49**:1078–1083.

Shilyansky J, Roberts EA, Superina RA. Distal splenorenal shunts for the treatment of severe thrombocytopenia from portal hypertension in children. *J Gastrointest Surg* 1999;**3**:167–172.

Sinha PS, Stoker TA, Aston NO. Traumatic pseudocyst of the spleen. *J R Soc Med* 1999;**92**:450–452.

Targarona EM, Balague C, Cerdan G, *et al.* Hand-assisted laparoscopic splenectomy (HALS) in cases of splenomegaly: a comparison analysis with conventional laparoscopic splenectomy. *Surg Endosc* 2002;**16**:426–430.

Touloukian RJ, Maharaj A, Ghoussoub R, Reyes M. Partial decapsulation of splenic epithelial cysts: studies on etiology and outcome. *J Pediatr Surg* 1997;**32**:272–274.

Tung CC, Chen FC, Lo CJ. Splenic abscess: an easily overlooked disease? *Am J Surg* 2006;**72**:322–325.

Ulhaci N, Meteoglu I, Kacar F, Ozbas S. Abscess of the spleen. *Pathol Oncol Res* 2004;**10**:234–236.

Wadsworth DT, Newman B, Abramson SJ, *et al.* Splenic lymphangiomatosis in children. *Radiology* 1997;**202**:173–176.

CHAPTER 27

Disorders of the pancreas

SIR ALFRED CUSCHIERI AND ABDOOL RAHIM MOOSSA

Surgical anatomy of the pancreas

Durmen has summarized the anatomical relationship of the pancreas as follows: 'The pancreas cuddles the left kidney, tickles the spleen, hugs the duodenum, cradles the aorta, opposes the inferior vena cava, dallies with the right renal pedicle, hides behind the posterior parietal peritoneum of the lesser sac and wraps itself around the superior mesenteric vessels.'

The pancreas is relatively inaccessible and, without some dissection, very little of it can be seen or palpated even at laparotomy. For this reason, it is much more difficult to manage surgically than all other abdominal viscera. Its retroperitoneal location in the upper abdomen means that it is almost completely hidden by the stomach, transverse colon and mesocolon (Figure 27.1).

The pancreas derives its blood supply from numerous arteries arising from major branches of the coeliac and superior mesenteric arteries. A full understanding of the local vascular anatomy and its possible variations is essential for any surgeon operating on the pancreas. In the thin patient, a part of the head of the gland may be seen directly behind the peritoneum of the supracolic and right infracolic compartments, and the inferior border of the body and tail may be visualized from the left infracolic compartment at the root of the transverse mesocolon. These views are very limited and are usually obscured by mesocolic and omental fat. The neck of the pancreas may be palpated by a finger passed through the epiploic foramen and directed inferiorly.

In order to inspect and palpate the pancreas properly three surgical manoeuvres are necessary.

- The hepatic flexure of the colon is mobilized downwards and medially by dividing its attachments to the duodenum and anterior aspect of the pancreatic head. The peritoneum lateral to the second part of the duodenum is incised and the duodenum and pancreatic head are elevated by blunt dissection (Kocher manoeuvre) from the posterior parietal structures. In this way, the right kidney, right renal vein, inferior vena cava and the root of the left renal vein are exposed. The head of the pancreas, the duodenum and retroduodenal and pancreatic portions of the common bile duct can thus be palpated between thumb and finger.
- Limited visualization of the superior part of the body of the pancreas may be obtained by opening an avascular part of the lesser omentum and retracting the lesser curvature of the stomach inferiorly. This manoeuvre also brings the coeliac axis into view. The body of the pancreas can only be adequately visualized by widely opening the gastrocolic omentum, retracting the transverse colon and mesocolon inferiorly and the greater curvature of the stomach superiorly. By extending this opening to the right, into the pyloric region, the right gastroepiploic vessels can be divided near their origins and the anterior aspect of the neck of the pancreas can be visualized. Care must be taken to avoid damage to the middle colic vessels in this region. The opening can also be extended to the left and the gastrosplenic ligament with its contained short gastric vessels can be divided to permit complete visualization of the anterior surface of the tail of the pancreas at the splenic hilum.
- Downward and medial retraction of the dome of the spleen will tense the peritoneal leaf known as the lienorenal ligament and this can be divided to allow the spleen, splenic vessels and tail of the pancreas to be mobilized *en bloc*, allowing inspection of the posterior aspect of the body and tail of the pancreas and more careful palpation of the distal portion of the gland.

All three manoeuvres should be carried out safely, quickly and with little risk of damage to vital structures or troublesome bleeding. They will allow evaluation of all areas of the pancreas

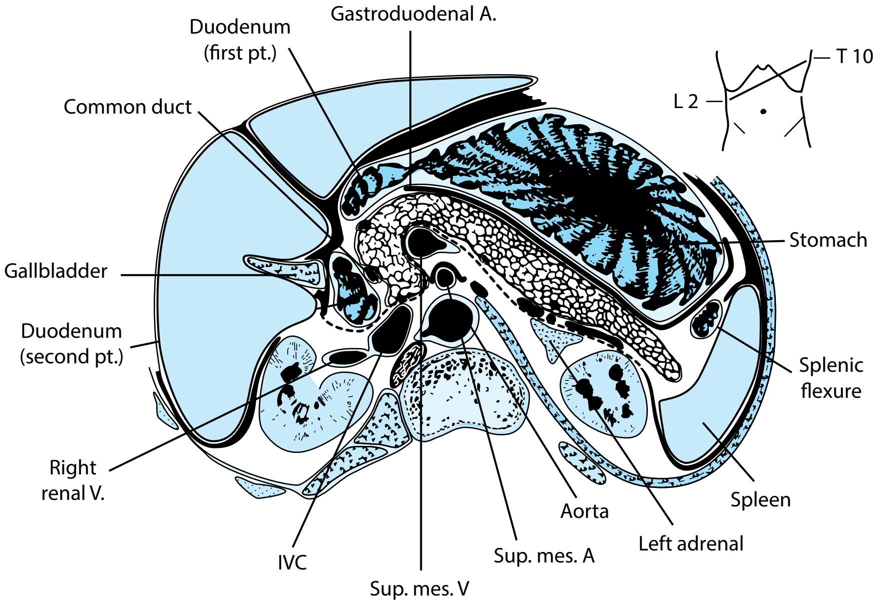


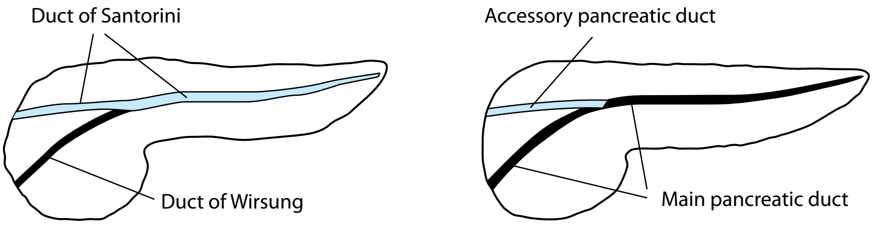
Figure 27.1 Oblique transverse cross-section of the upper abdomen viewed from below. Section passes through the long axis of the pancreas at approximately the levels indicated in the inset figure. The disposition and relations of structures shown approximate those seen in the oblique transverse scanning. A, artery; IVC, inferior vena cava; Sup. mes, superior mesenteric; V, vein.

except the region of the neck and uncinate process. Further dissection and mobilization is usually necessary to assess the resectability of a pancreatic tumour. For this, a detailed knowledge of the pancreas and peripancreatic vasculature and its variations is essential.

The coeliac axis and the superior mesenteric artery and their branches vary a great deal in both their site of origin and their direction. The same applies to the venous drainage of the foregut, its appendages and the midgut into the portal venous trunk. The most demanding part of a pancreatic resection is dissection of the neck and head of the gland from the superior mesenteric and portal veins and the uncinate process from the superior mesenteric artery. Hence, it was formerly considered mandatory to have a coeliac and a superior mesenteric arteriogram (as well as a venous phase) prior to planning any major pancreatic resection. Recent refinement in CT (multidetector helical CT scan) and MRI (3.0 T MRI) has rendered diagnostic pancreatic angiography largely obsolete.

As with the blood vessels, pancreaticobiliary ductal anatomy is very variable and the concept of a ‘normal duct anatomy’ is unsafe. The terminology which is widely applied to describe the main pancreatic ductal system is explained in Figure 27.2. The variations of the terminations of the main and accessory pancreatic ducts and their relationship to the lower end of the common bile duct are depicted in Figure 27.3.

The pancreas occupies a central position at a complex anatomical crossroads and its lymphatic drainage is radially disposed along several major routes, namely the coeliac, splenic, hepatic and superior mesenteric nodal basins. It is thus hard to design an adequate ‘cancer operation’ which, in an orderly manner, removes the primary tumour *en bloc* with primary, secondary and tertiary lymphatic nodal territories. Moreover, the intimate anatomical association of the pancreas with major vessels at once limits the extent of the procedure and dictates what must be removed. Thus, when a tumour of the pancreas spreads a short distance, it involves the portal vein, superior mesenteric artery or coeliac axis, and usually becomes incurable. Similarly, if the gland is removed in radical fashion, the need to excise the vessels and lymph nodes associated with it makes removal of the spleen, duodenum, gallbladder,



Embryological basis **Functional basis**

Figure 27.2 Terminology variously applied to describe the pancreatic ductal system. An understanding of ductal embryology, particularly with regard to the development of the more unusual variations, may be served by the terms given in (a). For clarity and practicality the terms given in (b) are preferred.

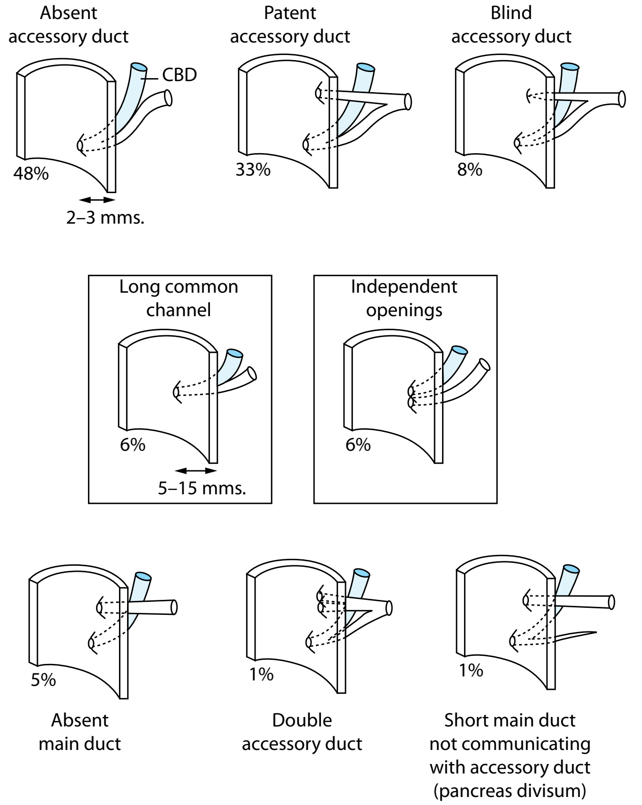


Figure 27.3 Variations of the main and accessory pancreatic ducts and their relationship to the common bile duct (CBD).

common bile duct, upper jejunum and part of the stomach necessary.

Where only part of the pancreas is excised or even if the gland is incised, safe management of any draining pancreatic juice becomes a matter of primary importance since enzymes, if allowed to accumulate in the peritoneal cavity, may cause local damage. A first principle of pancreatic surgery is the provision of adequate drainage. Similarly, the collection of serum, lymph and blood following a pancreatic resection needs to be drained. Second, absorbable suture material should not be used for ligature of major vessels, or as a suture material for anastomosis, or for closure of the abdomen during pancreatic surgery. Non-absorbable material, such as silk, nylon prolene, etc., is essential for safety.

Cellular composition and physiology of the exocrine pancreas

The exocrine pancreas consists of acinar and ductal systems which drain its secretions into the duodenum. The exocrine tissue accounts for 98% of the pancreas by weight. Under the influence of neural and hormonal controls, the exocrine pancreas secretes water and bicarbonate from the ductal system and enzymes from the acinar cells. The parasympathetic vagal fibres have ganglia in the interlobular septa, and postganglionic fibres are distributed to acinar cells and to smooth muscle cells in the ducts. The sympathetic fibres appear to be entirely distributed to the blood vessels and to be concerned solely with regulation of pancreatic blood flow rather than in the direct control of pancreatic secretion.

Fluid and electrolyte secretion from the pancreas is a ductal function and is an energy-requiring process. The cationic composition of pancreatic fluid is similar to that of plasma. Sodium and potassium concentrations are identical to those in plasma and are independent of flow. During states of fluid and electrolyte secretion, calcium appears to enter the ducts passively. However, under the influence of cholecystokinin (CCK), it appears to be actively secreted in parallel with enzyme secretion. Anionic secretion consists almost entirely of bicarbonate and chloride. The sum of concentrations of these two anions remains constant – a high chloride concentration occurs at low flow rates and chloride is replaced by bicarbonate as the flow increases.

Pancreatic enzyme secretion originates in the acinar cell and accounts for virtually all the protein (2–8 g/day in man) in pancreatic juice. Many of the enzymes are secreted in their inactive or zymogen forms together with inhibitors. This mechanism protects the pancreas from autodigestion by its own proteolytic enzymes. Enzyme activation begins after the zymogen enters the duodenum, where mucosal enterokinase cleaves trypsinogen into trypsin, leaving trypsin to activate the other enzymes. The large list of enzymes secreted in pancreatic juice includes amylase, lipase, cholesterol ester hydrolase, phospholecithinase A, trypsin, chymotrypsin A and B, elastase, carboxypeptidase A and B, collagenase, leucine aminopeptidase, ribonuclease and, undoubtedly, other enzymes for which a function has yet to be described.

Control of pancreatic exocrine secretion

Basal pancreatic secretion results either from an intrinsic autonomy of the gland or from a low level of activity of neurohormonal regulators. A complete neurohormonal control mechanism is at work and secretin, CCK and gastrin play the dominant roles. Secretin is released in response to duodenal acidification and there is also an increase in secretin release in man in response to alcohol and, to a lesser extent, after a meal. Secretin produces a secretion of fluid and electrolytes which is initiated within 30 seconds of an administered dose, the bicarbonate concentration of the fluid increasing as flow increases. It is now generally accepted that secretin is also a weak stimulant of enzyme secretion.

The stimulus for release of CCK appears to be entry of amino acids, fatty acids, hydrochloric acid and food into the duodenum. It causes an increase in the release of enzymes and a small increase in fluid and electrolyte output. In man, pancreatic secretion is initiated by CCK at a dose lower than that required for gallbladder contraction.

Gastrin has a varying effect on pancreatic secretion but, in man, it causes an increase in enzyme secretion. The action of glucagon and vasoactive intestinal peptide (VIP) on human pancreatic secretion has yet to be defined. Chymodenin appears to selectively induce chymotrypsin secretion. Somatostatin, pancreatic polypeptide (PP) and motilin have unidentified roles, but may act as a feedback control.

Hormonal and neural interaction

Combinations of two or more hormones have differing effects on the acinar and ductal cells. Secretin is a strong stimulant of fluid and electrolyte secretion and a weak stimulator of enzyme secretion; acting with CCK, however, marked augmentation occurs. These hormones have different receptor sites on the acinar cell and the site of interaction is probably intracellular. Such augmentation probably has an important physiological role since only small amounts of secretin are released into the circulation in response to a meal.

The interaction between the exocrine and endocrine cells of the pancreas is still being elucidated. Insulin is trophic to the peri-insular cells and a loss of insulin secretion in diabetes mellitus results in progressive damage to the acinar cell. The blood supply to the human pancreas first passes the islets and then forms a capillary network around the acinar cells, allowing maximal effects of islet hormones on the acinus. Thus, insulin, glucagon, somatostatin and other islet cell secretions may affect exocrine pancreatic secretion.

In man, vagal reflexes in response to gastric distension result in a juice rich in enzymes, and this effect is abolished by truncal vagotomy. It is likely that the role of vagal stimulation is a permissive one, allowing secretin and CCK to exert their full effect. Thus, atropine, by blocking acetylcholine, depresses the responsiveness of the acinar and ductal cells to CCK in animals.

Anatomy and physiology of the endocrine pancreas

The islets of Langerhans form the endocrine portion of the mammalian pancreas and consist of cells arranged in spherical or ovoid clusters which are well circumscribed and irregularly distributed throughout the gland. Although variable, the total number of islets in the adult human pancreas is estimated to be 1 million. The average weight of the entire pancreas in the adult approximates 100 g. The islets weigh about 1–2 g and thus form, at best, only about 2–3% of the weight of the whole gland.

Our knowledge of the cellular composition of the islets is still incomplete but the existence of four cell types has been largely accepted in mammalian islets based on ultrastructural characteristics of their secretory granules (Figure 27.4). These are the A-, B-, D- and F (D1)-cells.

A-cells form approximately 20% of the normal islet cell population, -cells 75%. In the various pathological conditions affecting the islets, including diabetes mellitus, the proportion of different cell types varies. A-cells (α-cells) synthesize, store and secrete glucagon. They are concentrated in the periphery of the islets. -cells (β-cells) synthesize, store and release insulin. Islet amyloid-associated polypeptide, or amylin, and pancreastatin are also formed in the pancreatic -cells. The physiological function of the former is unclear. However, pancreastatin has an inhibitory role on insulin secretion as well as glucagon and pancreatic acinar cell secretion in animal models. The role in humans is not well defined. -cells are concentrated in the centre of the islets. D-cells supply the tetradecapeptide somatostatin and also probably gastrin. They tend to be located in the periphery of the islets. F- also referred to as D1-cells are the rarest cell type encountered in normal islets and they are thought to be responsible for the secretion of PP. The F-cells constitute approximately 15–20% of cells within the islets of the ventral pancreas but are uncommonly found in the dorsal pancreatic islets. The number of F-cells probably increases with age and in association with injury to pancreatic tissue. An increase in the number of F-cells has been documented in such conditions as diabetes, pancreatitis, mucoviscidosis, haemochromatosis and various exocrine or endocrine pancreatic tumours.

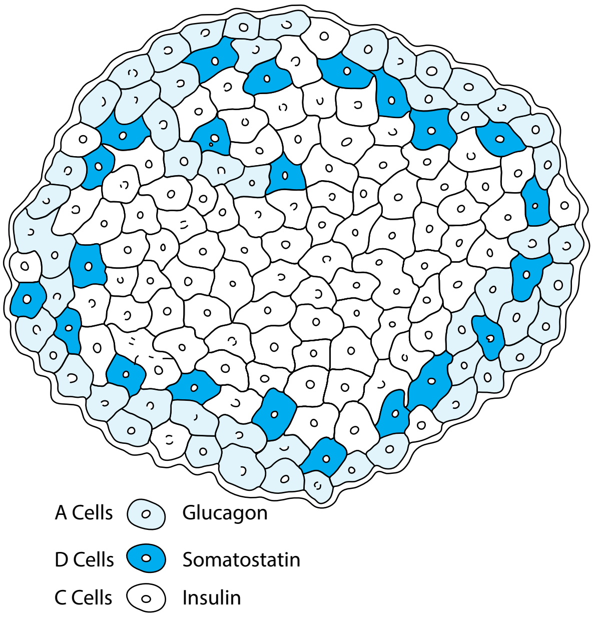


Figure 27.4 Schematic representation of an islet of Langerhans showing distribution of the three main cell types (From Orci, Unger, *Lancet* 1975; ii:1243, with permission.)

The pancreatic islets consist of cell types in close topographical relationship, which produce hormones with different but related actions. Gaps and tight junctions have been demonstrated between islet cells of the same as well as different types. This arrangement may play a role in co-coordinating the activity of the different cell types, e.g. somatostatin-containing cells are topographically closely related to B- and A-cells. As somatostatin inhibits insulin and glycogen release, it is possible that it serves a paracrine function by regulating the local (within the islet) release of these hormones from the appropriate cells. It is likely that the islets function as a well-integrated unit and further elucidation of such mechanisms will enhance the understanding of various pathological conditions arising in this endocrine organ.

Insulin

Insulin is secreted only by the -cells whereas the other hormones are also secreted by the gastrointestinal mucosa, and somatostatin is also found in the brain. Both insulin and glucagon are important in the regulation of carbohydrate, protein and lipid metabolism with insulin functioning as an anabolic hormone (increases the storage of glucose, fatty acids and amino acids in cells and tissues), and glucagon antagonizing insulin by its function as a catabolic hormone mobilizes glucose, fatty acids and amino acids from stores into the bloodstream.

Insulin has a molecular weight of 5800 and consists of two peptide chains A and B which are connected by two disulphide bridges (Figure 27.5). In man, the A-chain has 21 amino acids and the B-chain is composed of 30 amino acids. There are structural variations from species to species. It is known that insulin is formed in a precursor form, proinsulin, which comprises the insulin A and B chains linked by a polypeptide segment consisting of 30–35 amino acids. This connecting peptide (C-peptide) helps in the formation of the native structure of the insulin molecule by ensuring the correct pairing of the cysteine residues during formation of the disulphide linkages between the A- and B-chains. Proteolytic cleavage of proinsulin by enzymes including kallikrein in the secretory granules results in the separation of insulin from the C-peptide. In this conversion, a C-peptide chain is removed from the proinsulin molecule producing the disulphide-connected A- and B-chains of insulin.

These two products are thus released in equimolar amounts from the -cells. Proinsulin-like components (PLCs) constitute 15% of the immunoreactive insulin concentration. The kidney is an important site of proinsulin degradation; hence levels of PLCs are markedly elevated in chronic renal failure. Raised serum proinsulin concentrations have been found in both benign and malignant insulinomas. Neither insulin nor insulin antibodies interfere in the immunoassay for the estimation of C-peptide concentration. Since insulin and C-peptide are formed in equimolar amounts following cleavage of proinsulin, the circulating levels of C-peptide provide a measure of -cell secretory activity, especially in the presence of circulating insulin antibodies produced in response to exogenous insulin. Thus it gives an indication of endogenous insulin production from the β-cell. C-peptide measurement therefore has a vital role in the diagnosis of surreptitious injection of insulin and in the follow-up evaluation of patients who have undergone total or near-total pancreatectomy for nesidioblastosis or a malignant insulinoma. Following resection of the latter, significant C-peptide levels indicate residual pancreatic tissue or the presence of functioning metastases. Recurrence may be heralded by a rise in C-peptide levels.

The main stimulus to insulin secretion is an increase in glucose levels. The insulin will thus cause a fall in glucose concentration and, by a negative feedback effect, this leads to a decrease in insulin release. Other stimuli for insulin release are the amino acids arginine and lysine, glucagon, growth hormone, cortisol, placental lactogen and the sex hormones (especially oestrogens). Glucagon releases insulin both by direct stimulatory effect on the -cells and also indirectly by mobilizing glucose from the liver. The remaining hormones mentioned above act by causing resistance to the actions of insulin and thus generate a compensatory increase in insulin secretion. A number of gastrointestinal hormones including gastrin, secretin, gut glucagon, CCK-PZ (pancreozymin) and gastric inhibitory polypeptide may also be stimulants for insulin secretion. Insulin release is also enhanced by free fatty acids to prevent ketoacidosis which would otherwise occur during fasting in normal individuals.

Vagal stimulation is also a potent stimulus to insulin release, although no impairment in insulin release can be demonstrated following vagotomy. Adrenergic β-receptors also stimulate insulin release, but epinephrine (adrenaline) and norepinephrine (noradrenaline), acting via α-receptors, have an inhibitory effect on insulin secretion. The ventromedial nucleus of the hypothalamus is believed to be important in the cephalic phase of insulin release. Finally, insulin secretion is influenced by various drugs. Tolbutamide and chlorpropamide have a stimulatory effect while alloxan and streptozotocin impair insulin secretion by directly damaging the -cells.

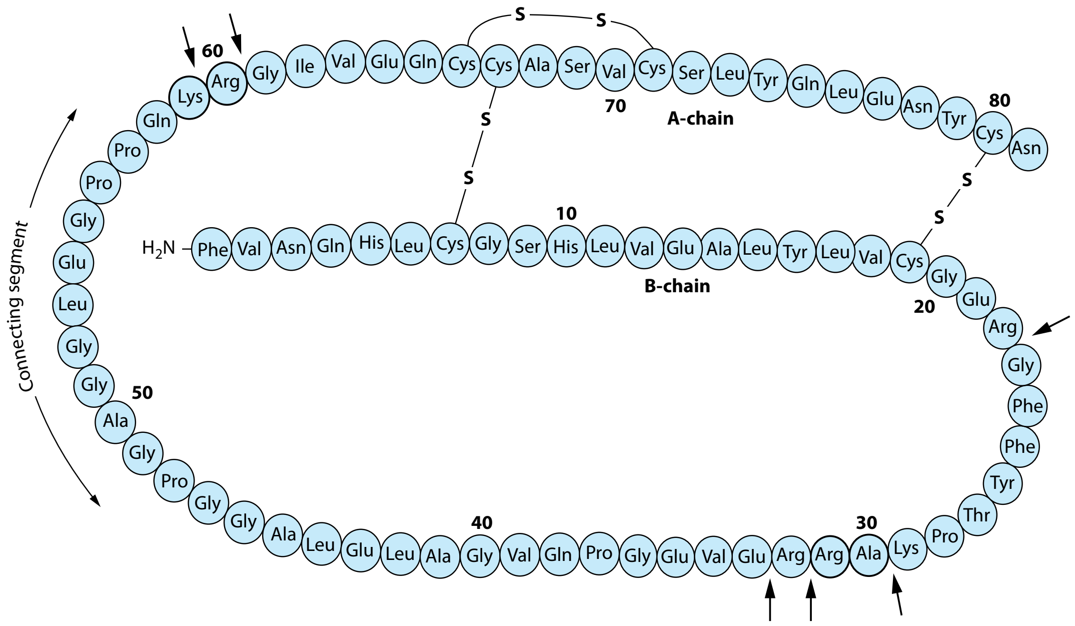

S
S
S
S
S
S
60
70
80
10
20
30
40
50
A-chain
B-chain
Connecting segment
H_2N –
Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Ala
Arg Arg Glu Val Glu Gly Pro Gln Val Gly Ala Leu Leu Glu Leu Ala Gly Gly Pro Gly Ala Gly Gly Leu Glu Gly Pro Pro Gln Lys Arg
Gly Ile Val Glu Gln Cys Cys Ala Ser Val Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn

Figure 27.5 Covalent structure of bovine proinsulin. Arrows indicate sites of cleavage by trypsin.

Secretion and actions of insulin

Insulin secretion is pulsatile (occurs in bursts), regulated by a variety of stimulatory factors, most of them related to glucose metabolism and the effects of cyclic adenosine monophosphate (cAMP). Insulin secretion is stimulated by high blood glucose levels and reduced by hypoglycaemia. Other stimulants include several amino acids, intestinal hormones, acetylcholine (parasympathetic stimulation). Inhibitory factors for insulin secretion include somatostatin and norepinephrine (noradrenaline) (sympathetic stimulation). Once released in the circulation, insulin is degraded within minutes in the liver and kidneys. C-peptide and kallikrein are also present in the circulation, having been secreted with insulin. Antibodies to components of islet cells develop in most patients with insulin-dependent diabetes.

The major role of insulin is to promote entry of glucose and amino acids in cells. Glucose enters cells by facilitated diffusion along a concentration gradient created by low intracellular free glucose and by the availability of a specific glucose transporter. Insulin exerts its action by binding with specific membrane receptors found on the surface of cell membranes. The insulin receptor, a tetramer, is made up of two α and two β-glycoprotein subunits. The β-subunit is a protein kinase that catalyses the phosphorylation of proteins. This phosphorylation increases the activity and number of glucose transporters (protein carriers of glucose). As the intracellular glucose concentration is low, some glucose enters the cell even in the absence of insulin. With insulin, however, the rate of glucose entry is considerably accelerated by the process of facilitated diffusion mediated by the transporters. After the insulin receptor complex enters the cell it is transported to the lysosomes where it is cleaved, and the receptor recycled. Raised circulating levels of insulin reduce the number of receptors (downregulation), whereas decreased insulin levels has the opposite effect – upregulation of the number of receptors. The density of receptors per cell is increased in starvation and decreased in obesity and acromegaly. The receptor affinity is decreased by excess glucocorticoids.

The major actions of insulin are:

- facilitation of glucose transport through cell membranes
- stimulation of the enzyme system involved in conversion of glucose to glycogen in liver and muscle
- depression of gluconeogenesis in liver and muscle cells
- regulation of lipogenesis in liver and adipose tissue
- promotion of protein synthesis and growth.

In the liver, insulin promotes glycogen deposition through glycogen synthetase and thus increases in glucose uptake. Insulin suppresses the synthesis of key gluconeogenic enzymes and induces the synthesis of key glycolytic enzymes such as glucokinase. Insulin likewise increases the activity of enzymes involved in lipogenesis.

Glucagon

Glucagon is a secretory product of the A-cell and is a linear peptide composed of 29 amino acids with a molecular weight of approximately 3500. Gel filtration of acid alcohol extracts of pancreatic tissue has revealed two peaks of immunoreactivity, one with a molecular weight in excess of 9000 (believed to represent proglucagon) and another which is a globulin-sized fraction that has been referred to as big plasma glucagon, which may be a precursor of glucagon or simply glucagon bound to a larger protein.

Glucagon has catabolic actions. Hypoglycaemia (a fall below 90 mg/dL) produces a rise in plasma glucagon concentration, and an increase in glucose concentration leads to a drop in glucagon levels. Glucagon activates adenyl cyclase which increases hepatic cAMP to initiate breakdown of glycogen into glucose (glycogenolysis). It also inhibits the process of glycogenesis. It increases glucose formation from non-glucose precursors, e.g. glycerol, lactate and amino acids of the glycogenic type. Glucagon also enhances the breakdown of fat into free fatty acids and glycerol. It stimulates gluconeogenesis and, thus, its infusion lowers plasma amino acids. Finally, glucagon inhibits protein synthesis. There is evidence that glucagon, through its gluconeogenic, ketogenic and lipolytic effects, and not lack of insulin alone, is partly responsible for the development of fulminant diabetic ketoacidosis in man.

The inter-relationship between insulin and glucagon is a complex one which is only partly understood. In general, an inverse relationship exists between insulin and glucagon levels. When glucose is needed, insulin levels fall and glucagon levels rise, producing an increased hepatic glucose production. The reverse is true in states of hyperglycaemia. Following a protein meal, a parallel change is observed in the levels of insulin and glucagon. The rise in glucagon level prevents the hypoglycaemia that would result from enhanced insulin secretion alone by amino acids. CCK-PZ, which stimulates insulin secretion, is also believed to facilitate glucagon secretion following the stimulus of a protein meal and is responsible for the buffer mechanism against hypoglycaemia arising from ingestion of protein.

Somatostatin

This tetradecapeptide is secreted by the D-cell. It inhibits the release of growth hormone from the anterior pituitary and was first isolated from the hypothalamus. Its other actions include inhibition of insulin and glucagon secretion, gastrin secretion, acid and pepsin secretion from the stomach, as well as the release of pancreatic enzymes. It suppresses intestinal motility and contraction of the gallbladder. It also has a suppressive effect on glucose uptake from the gut and on appetite and may play a role in nutrient homeostasis. Infusion of somatostatin diminishes splanchnic blood flow. In view of its inhibitory effect on the exocrine pancreas, the long-acting somatostatin analogue is used in the treatment of patients with pancreatic fistula and acute pancreatitis. Somatostatin has been isolated from a variety of tissues and organs, including the gastrointestinal tract and pancreatic islets. Its exact and full spectrum of function is as yet unknown but it is thought to be involved locally in the regulation of the secretion of the other pancreatic hormones. Somatostatin may act as a neurotransmitter in the brain (hypothalamus) and spinal cord.

Pancreatic polypeptide

PP is a member of the neuropeptide Y family, the structure of which includes a carboxy-terminal tyrosine amide and a chain of 36 amino acids. The family includes PP, peptideYY (PYY) and neuropeptide Y (NPY). They are found in different locations throughout the gastrointestinal tract and in the nervous system; and despite structural similarity, they have different biological actions. PP is expressed in endocrine cells of the gut and pancreas, PYY is located in enteroendocrine mucosal cells of the ileum and colon and nerves of the enteric nervous system, and NPY acts as a neurotransmitter in both central and peripheral nervous systems. PYY occurs in high concentration in the ileum and colon in two different cell types: L cells of the ileum where it is collocalized with enteroglucagon and H-cells of the colon and rectum. All three peptides exert their biological action by binding to various G protein-coupled receptors known as 'Y' receptors. They signal through the GTP-binding proteins Gi/Go and inhibit adenyl cyclase. The PP family of peptides functions as endocrine, paracrine, and neurocrine transmitters. PYY and PP are hormones acting in both a paracrine and an endocrine fashion. Following a meal, PP is released into the blood from PP cells of the pancreas via vagal cholinergic stimulation. PP blood concentrations also fluctuate with the myoelectric activity of the gastrointestinal tract. In humans PYY is released primarily in response to fatty meals. All disorders causing malabsorption increase blood levels of PYY. Additionally, surgical interventions which increase the delivery of food to the distal ileum and colon (small bowel resection, vertical band gastroplasty, Roux-en-Y gastric bypass, jejunoileal bypass) induce elevation of the circulating PYY.

PP is synthesized by the pancreatic islet F-cells with a 29 amino acid sequence and a 66 amino acid prohormone that contains the 36 amino acid PP hormone, and a 20 amino acid icosapeptide of unknown function and C-terminal prosequence. It is released into the circulation following a meal (protein and fat). PP slows gastric emptying and insulin secretion and is thought to inhibit further intake of food. As most of the PP is in the pancreas, it becomes almost undetectable in pancreatectomized patients. Its secretion is believed to be largely due to vagal cholinergic stimulation. PP is thought to play a role in promoting intraislet homeostasis and to modulate gastrointestinal function.

In pharmacological doses, PP inhibits pancreatic secretion caused by CCK-PZ and secretin, stimulates gastric emptying and intestinal transit, inhibits relaxation of the gallbladder and increases the tone of the choledochal sphincter. It has been suggested as a possible causative agent in the watery diarrhoea–hypokalaemia–achlorhydria (WDHA) syndrome. Significant elevated levels of PP are found in both maturity-onset and juvenile diabetics. This observation has also been extended to patients with chronic pancreatitis. Many endocrine tumours of the pancreas and their metastases have also been found to contain numerous PP cells and also to have a high tumour content of PP.

Methods of investigating the pancreas

Because of its deep-seated and inaccessible location, the pancreas is a difficult organ to investigate and to visualize. A precise diagnosis of pancreatic disease is often only possible through the use of a wide battery of tests. The results of such tests should be viewed in the light of the clinical information since all available procedures may not yield concordant data.

Procedures which are employed in the investigation of patients with suspected pancreatic disorders may be classified into five groups:

- Procedures which outline the gland to delineate enlargement, masses, irregularities in contour and calcification. These include:
  - indirect imaging of the pancreas:
    - standard radiological studies to visualize the effect of the pancreas on adjacent organs such as stomach, duodenum, small bowel, transverse colon and bile duct; these studies are largely outdated and have been replaced by other more modern imaging techniques
  - direct imaging techniques to visualize
    - the pancreatic parenchyma, e.g. ultrasonography, CT, magnetic resonance cholangiopancreatography (MRCP), endoscopic ultrasound (EUS).
    - the pancreatic duct system – endoscopic retrograde cholangiopancreatography (ERCP).
    - the pancreatic and peripancreatic vasculature – angiography, contrast CT scan, MRI.
- Procedures to define pancreatic exocrine function:
  - faecal fat excretion
  - pancreatic function tests.
- Procedures to define pancreatic endocrine function:
  - measurement of fasting blood levels of glucose and/or hormones which are secreted by the pancreatic islets under normal and/or abnormal conditions; these include serum insulin, proinsulin, C-peptide, glucagon, somatostatin and gastrin
  - provocative tests to measure the serum level of the above substances if the fasting levels are not conclusive:
    - calcium infusion test (insulinoma and gastrinoma)
    - tolbutamide tolerance test (insulinoma)
    - glucagon test (insulinoma)
    - insulin suppression test (insulinoma)
    - secretin test (gastrinoma).
- Analysis of serum for markers of pancreatic disease:
  - enzymes, such as amylase, lipase, trypsin, ribonuclease
  - tumour-associated antigens, such as carcinoembryonic antigen (CEA), pancreatic oncofetal antigen (POA) and carbohydrate antigen (CA) 19-9.
- Pancreatic biopsy and cytology:
  - percutaneous fine-needle aspiration (FNA) cytology using ultrasonography or CT for guidance
  - endoscopic transgastric or transduodenal needle aspiration cytology or brush cytology
  - laparoscopic visualization and direct vision biopsy or aspiration cytology of the pancreas
  - operative visualization, palpation and biopsy of the pancreas.

Standard radiological investigations

Standard or routine radiological investigations rely on the detection of anatomical abnormalities by the displacement or distortion of adjacent viscera, such as the pancreas. Maximum information may be obtained from each type of examination if the radiologist involved is alerted to the possibility of pancreatic disease prior to the actual procedure. The main value of the various investigations to be described is that they provide important information in a clinical setting suggestive of disease in the upper abdomen. They should be employed principally to identify and/or to exclude common disorders such as peptic ulcer, gallstones, hiatus hernia, gastric cancer and colon cancer. They may, however, show a wide range of abnormalities suggestive of pancreatic disease. Two common pitfalls need to be emphasized: (1) the presence of a common benign disorder, such as hiatus hernia or gallstones does not preclude the simultaneous presence of pancreatic disease; (2) normal routine radiological studies do *not* necessarily signify the absence of pancreatic disease.

A *plain radiograph of the abdomen* may show changes suggestive of *pancreatic disease*, the most important of which is pancreatic calcification. Radiating ‘sun-burst’ calcification is pathognomonic of cystadenomas or cystadenocarcinomas of the pancreas. Numerous other lesions of the pancreas may also show calcifications. These include chronic pancreatitis, with or without pancreaticolithiasis, lymphangiomas, haemangiomas and, occasionally, mucin-secreting adenocarcinoma or islet cell carcinoma. In hereditary pancreatitis, the incidence of pancreatic calcification is higher than in other types of pancreatitis and the incidence of pancreatic malignancy is also increased. In general 2–4% of all patients with pancreatic calcification have a coexisting pancreatic carcinoma but, conversely, over 95% of all patients with pancreatic calcification will have benign disease.

Contrast studies of the gastrointestinal tract

Pancreatic disease may reflect changes in the oesophagus, stomach, small bowel and colon. A pancreatic pseudocyst may present in the posterior mediastinum and compress the oesophagus. Occasionally, a tumour arising in the tail of the pancreas will also deform and involve the distal oesophagus. Alternatively, metastatic lymphadenopathy in the posterior mediastinum may also occlude the oesophagus. Gastric and/or oesophageal varices occasionally accompany pancreatitis or any type of pancreatic cancer as a result of splenic vein occlusion.

Radiological examination of the stomach and duodenum

Changes produced in the stomach and duodenum can be considered under the headings of extrinsic compression and indentation, rugal and mucosal abnormalities, enlarged retrogastric space and gastric varices. If a double contrast examination is performed, some of the more subtle mucosal changes may be seen to better advantage. Motility changes of the stomach, particularly the antrum, may be seen when the posterior stomach wall is invaded by an inflammatory process or a malignant disease of the pancreas.

Pancreatic disease may reflect on the duodenum in a number of different ways and these may be seen on contrast radiography as pressure defects, abnormalities of duodenal fold pattern, widening of the duodenal C-loop, displacement of the angle of Trietz, postbulbar ulceration of the duodenum (Zollinger–Ellison syndrome), disorders of duodenal motility under fluoroscopy, enlargement of the ampulla of Vater and duodenobiliary reflux. It should be emphasized that by the time the presence of a pancreatic carcinoma is reflected by diagnosable changes on contrast radiology of the gastrointestinal tract, the lesion is advanced and incurable.

Radiological examination of the small bowel

Mass lesions of the pancreas may produce displacement of the duodenojejunal area and of the small bowel. These appearances are usually seen with pancreatic pseudocysts and large tumours. Chronic pancreatic disease associated with exocrine insufficiency and steatorrhoea may show a malabsorption pattern of the small bowel with thickened, clubbed or effaced folds. The classic appearance is seen in cystic fibrosis. Patients with the Zollinger–Ellison syndrome may show thickening of the folds in the duodenum and proximal jejunum and hypersecretion of fluid with dilution of the barium. Multiple peptic ulcerations may also be seen.

Radiological examination of the colon

In pancreatitis, characteristic changes have been described in the transverse colon and the region of the splenic flexure. These include ileus of the transverse colon (colon cut-off sign), displacement of the transverse colon and, sometimes, colonic strictures, fistulas and necrosis. Some of these appearances may simulate carcinoma of the colon. Rarely, a pancreatic pseudocyst may present in an unusual location and has even been described presenting as a presacral mass indenting the rectum. Intraperitoneal seeding from metastatic pancreatic carcinoma may cause indentation of any part of the colon.

Cholangiography and isotope biliary excretion scintigraphy

Patients with pancreatic disease (cancer and/or pancreatitis) may present with biliary tract obstruction and jaundice. If the serum bilirubin exceeds 51 μmol/L (3 mg/100 mL), then the outdated oral or intravenous cholangiography and Tc-HIDA scan no longer plays a role in the investigation. If ultrasonography or CT shows dilated biliary radicles, then percutaneous transhepatic cholangiography has a place in the investigation of biliary obstruction. However, when the biliary obstruction is suspected to be in the pancreatic region, ERCP or MRCP are the preferred methods of visualizing the biliary tree since they can provide valuable additional information.

Direct imaging of the pancreas

Ultrasonography

Ultrasonography is probably the most important advance in pancreatic investigation of the past three decades. The following features are indicators of pancreatic disease on ultrasonography.

- Diffuse or localized enlargement of the gland may be due to inflammation, tumour or pseudocysts. Atrophy of the entire pancreas usually indicates chronic pancreatitis.
- Alteration of the texture or 'pattern of internal echoes' within the pancreas may provide a *subjective* impression of pancreatic disease.
- Indirect signs outside the pancreas but readily traced to the pancreatic area are dilatation of the biliary system and displacement of the vessels adjacent to the pancreas.
- Abnormal dilatation of the pancreatic duct may be the result of cancer or chronic pancreatitis.
- Metastases and ascites are also indirect signs of pancreatic disease which by themselves do not necessarily imply the presence of an abnormal pancreas.

The distinction between pancreatic cancer, chronic pancreatitis and a variety of other pancreatic tumours can be difficult without knowledge of the clinical situation. External abdominal ultrasonography is relatively inexpensive, non-invasive and free of radiation hazards. It can be repeated as often as necessary but requires good equipment and expert staff for its performance and its interpretation. In spite of this, technical failure which prevents proper visualization of the whole pancreas occurs in about 10–15% of patients and can be attributed to obesity, gastrointestinal gas and previous operations in the upper abdomen as well as massive ascites.

Computed tomography (Figures 27.6 and 27.7)

CT scanning obtained by the current generation of multidetector CT machines (helical CT) is highly successful in demonstrating the pancreas and allows complete visualization of between 93% and 100% of all glands. The most valuable sign of disease shown by CT scan is the presence of localized or diffuse enlargement of the gland. As with ultrasonography, this finding is non-specific as to the type of disease. The pancreas may be abnormal without recognizable enlargement and, conversely, normal variants of size exist which may suggest abnormalities when none is present. An abnormally small pancreas may be a sign of chronic pancreatitis. The lower limit of normal is said to be one-half of a vertebral body width for the head of the pancreas and one-third for the body of the gland. Nevertheless, when the size of the gland falls below these standards, the pancreas is often actually normal. Calcifications are readily identified by CT scanning. Other signs of pancreatic disease include dilatation of the hepatobiliary tree, dilatation of the pancreatic duct and liver metastases.

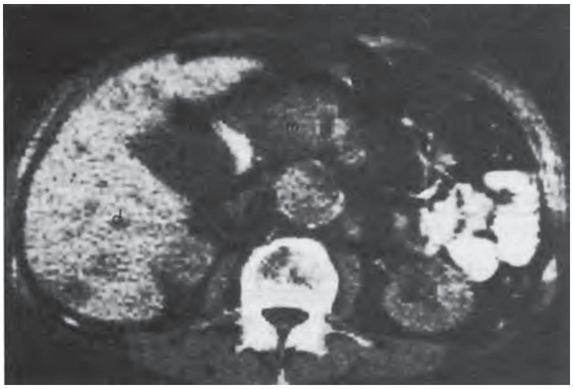


Figure 27.6 CT scan showing typical appearance of carcinoma of the pancreas, consisting of a mass (m) in the head of the pancreas, a distended gallbladder (g) and dilated intrahepatic ducts (d). Incidental note is made of an aortic aneurysm which was better seen in more caudal sections.

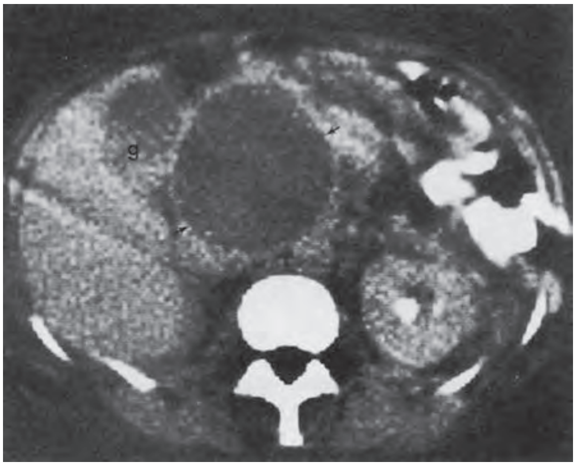


Figure 27.7 Scan shows typical pseudocyst (arrows) near the head of pancreas. The gallbladder (g) is also visualized.

The accuracy of ultrasonography and CT scan in detecting abnormalities in the pancreas varies. The reported figures are dependent on the type and spectrum of diseases included in the population studied, on the nature of the equipment employed, and on the skill of the individuals involved in the study. Chronic pancreatitis produces more problems for ultrasonic detection – only about 50% of cases of chronic pancreatitis may show changes detectable by ultrasound. One major difference between ultrasonography and CT scan is the frequency with which the gland can be seen. Ultrasonography has a non-visualization rate that averages 15–20% when conventional techniques are used. What is more subtle, and not known, is the frequency with which significant portions of the pancreas are hidden, even though the investigator thinks the entire gland is seen. On the other hand, failure rates for CT are significantly lower and range from 0% to 10%. CT scan has a higher chance of visualizing the gland when ascites or extreme obesity is present. On the other hand, when the patient cannot remain motionless or suspend respiration during the scanning, CT study is usually inadequate but the ultrasound examination may produce diagnostic results. Recognition of pancreatic calcification and small intrapancreatic pseudocysts is clearer with CT scan and this facilitates diagnosis of pancreatitis. Finally, detection of gas and of the thicker walls of pancreatic abscesses is easier with a CT scan. CT examinations are two to four times more expensive than pancreatic ultrasound. When serial or multiple studies are required, the small but potential risk of radiation exposure has to be taken into account with the CT scan.

Contrast-enhanced helical CT is nowadays extensively used for the diagnosis and staging of pancreatic neoplasms, especially pancreatic ductal adenocarcinoma (PDAC) (the commonest cancer). The non-ionic contrast agent (300 mg of iodine/mL) used is injected intravenously in doses ranging from 120 to 150 mL at rates ranging from 1.5 to 5 mL/s with a time delay between start of the injection and scanning of 25–70 seconds. The pancreatic phase during which the pancreatic parenchyma is most enhanced usually occurs following infusion of 150 mL of contrast injected at 3 mL/s but this period varies with the injected rate. As PDACs are hypovascular, the detection rate especially of small cancers depends on maximum enhancement of the pancreatic parenchyma. A study has confirmed that higher doses and fast injection rates produce the best enhancement and thus the highest diagnostic yield.

Endoscopic retrograde cholangiopancreatography

Endoscopic diagnosis at ERCP

ERCP endoscopy provides direct visual observation of the oesophagus, stomach, duodenum and ampulla of Vater. Inflammatory or neoplastic lesions visualized in any part of the upper gastrointestinal tract may entirely explain the clinical picture.

Tissue diagnosis at ERCP

Conventional endoscopic biopsy with forceps, brush cytology, aspiration of pancreatic juice for cytology or direct transduodenal needle aspiration cytology may be possible and provides valuable information if positive for malignant disease. In addition, aspirated pancreatic juice may be assayed for research into tumour markers such as CEA and POA.

Radiological diagnosis at ERCP (Figures 27.8–27.10)

Retrograde pancreatogram and/or retrograde cholangiogram often provide information indicative of the presence of tumour, inflammatory disorders or stone disease.

The main disadvantage of ERCP is that it requires the combined team effort of an endoscopist, an assistant who is thoroughly familiar with the equipment and technique (including disinfection procedures and the handling of specimens), a radiology technician and, ideally, a radiologist. It is time-consuming and expensive. It is an invasive procedure which carries a small but definite risk of complications. ERCP is unsuitable for screening patients with unsuspected pancreatic disease.

The complications of ERCP include the following.

- Injection pancreatitis – this can be prevented by careful injection of the medium into the pancreatic duct without too much pressure and without attempting to obtain an acinar phase of the pancreatogram.

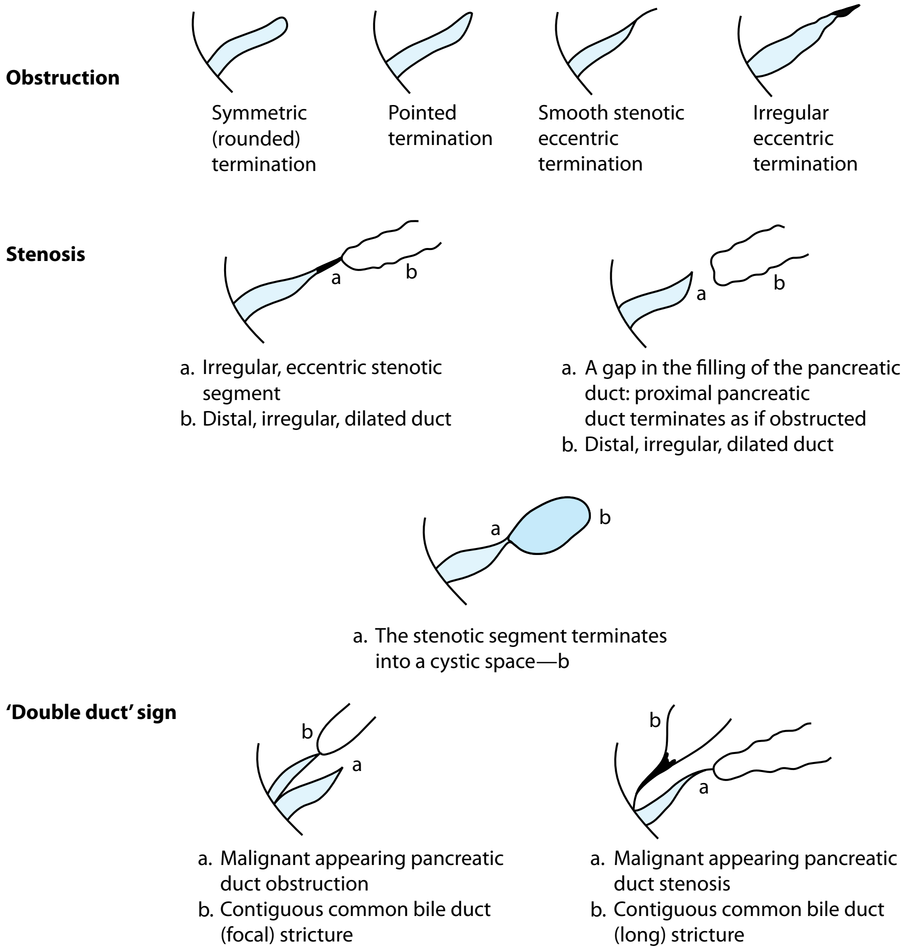

Obstruction
Symmetric (rounded) termination
Pointed termination
Smooth stenotic eccentric termination
Irregular eccentric termination
Stenosis
a
b
a. Irregular, eccentric stenotic segment
b. Distal, irregular, dilated duct
a
b
a. A gap in the filling of the pancreatic duct: proximal pancreatic duct terminates as if obstructed
b. Distal, irregular, dilated duct
a
b
a. The stenotic segment terminates into a cystic space—b
'Double duct' sign
b
a
a. Malignant appearing pancreatic duct obstruction
b. Contiguous common bile duct (focal) stricture
b
a
a. Malignant appearing pancreatic duct stenosis
b. Contiguous common bile duct (long) stricture

Figure 27.8 Diagram of typical radiographic appearances of the main pancreatic duct as demonstrated by endoscopic retrograde cholangiopancreatography in patients with cancer of the head of pancreas.

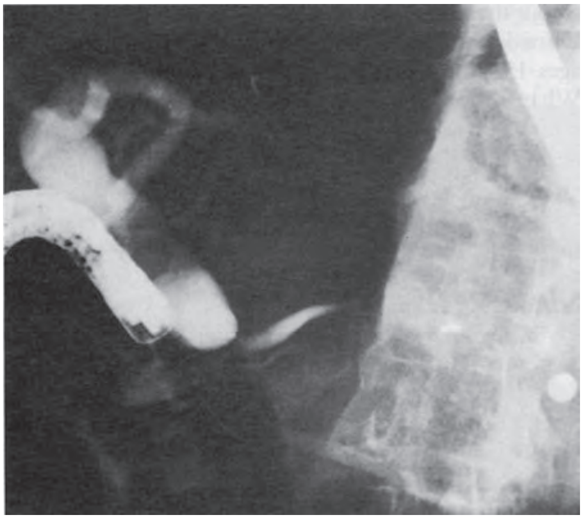


Figure 27.9 Endoscopic retrograde cholangiopancreatography showing 'rat tail' appearance due to obstruction of the main pancreatic duct by a carcinoma. The common bile duct is dilated. The 'double duct' sign is highly suggestive of unresectablility of the cancer.

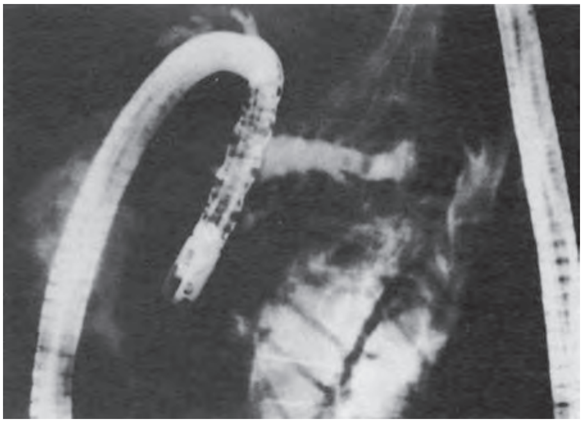


Figure 27.10 Endoscopic retrograde cholangiopancreatography showing grossly dilated main pancreatic duct due to chronic pancreatitis.

- Sepsis – pancreatic and biliary sepsis can be induced or precipitated by ERCP. Whether the sepsis is of endogenous or exogenous origin is a debatable and academic point. Suffice to say that ERCP sepsis is most frequently related to injection into an obstructed pancreatic duct, a pseudocyst or an obstructed common bile duct.

ERCP sepsis can be largely prevented or abolished if the following precautions are taken:

- disinfection of the endoscope tip and the cannula with 2% glutaraldehyde
- prophylactic systemic antibiotics intravenously for 24–48 hours
- addition of antibiotics (chloramphenicol or gentamicin) to the contrast material
- the patient should be operated on within 24 hours of ERCP if a pseudocyst or an obstructed duct is demonstrated.

Other rare complications of ERCP are drug reactions, instrumental injury to the upper gastrointestinal tract and aspiration pneumonia which is associated with the amount of sedatives given to the patient.

Angiography

Percutaneous transfemoral catheterization for coeliac and superior mesenteric angiography is, occasionally, a valuable method of studying the vasculature of the pancreas. With improvement in and refinement of catheter design, various techniques of superselective catheterization may be enhanced by magnification radiography and, in some instances, by the intra-arterial injection of various drugs which improve the visualization (pharmacoangiography) of pancreatic vessels. Vasodilators (bradykinin, tolazolin) and/or vasoconstrictors (epinephrine, norepinephrine, angiotensin) and hormones (secretin and pancreozymin) have all been tried.

The single most useful and reliable angiographic sign of a malignant tumour of the pancreas is arterial encasement. Encasement is seen as a narrowing and/or irregularity of a vessel and is caused by invasion of the vessel by tumour or its compression by surrounding tissues. Arterial encasement may be irregular or smooth or may be serrated or serpiginous. The smooth encasement is much less specific for cancer and may be seen in pancreatitis. Large artery (splenic, hepatic, superior mesenteric, left gastric) encasement is highly suggestive of an unresectable tumour. Small artery encasement is a term applied when more distal branches supplying the pancreas are involved. In connection with the assessment of tumour resectability, the gastroduodenal artery is considered to be borderline between the two groups. A second angiographic sign of unquestionable value in the diagnosis of pancreatic cancer is the presence of major venous involvement. An adequate venous phase angiogram frequently reveals obstruction and narrowing or deformity of the veins. However, non-visualization of the splenic or portal vein alone, although suggestive, is not diagnostic of venous obstruction. Arterial occlusion or arterial displacement may also be caused by pancreatic tumours. Neovascularity is only *rarely* seen in pancreatic carcinoma because the cancer is avascular. However, hyperaemia or increased vascularity in the region of the pancreas may be seen in cases of pancreatic cancer and is usually due to secondary inflammatory changes around the tumour. Finally, an angiogram may occasionally disclose the presence of hepatic metastases, thereby suggesting the malignant nature of a pancreatic abnormality.

Islet cell tumours are typically vascular. A fine network of small vessels may be seen in the early arterial phase and, occasionally, feeding arteries and draining veins may be identified. With larger tumours, displacement of neighbouring arteries may be seen and larger, irregular 'tumour vessels' have been reported. It is difficult to correlate angiographic visualization of islet cell tumours with their malignant potential unless hepatic metastases are obviously seen or encasement (or occlusion) of arteries is a striking feature. When angiography does demonstrate a solitary islet cell tumour, the surgeon must be aware that there may be a second lesion not shown angiographically. When angiography shows multiple islet cell tumours, the likelihood of there being others not shown is very real. In this respect, angiography does not remove the surgeon's obligation to carry out a complete exploration of the whole pancreas.

A major value of angiography is to delineate variations in the foregut vasculature, especially the hepatic blood supply. The desirability of obtaining angiographic studies before embarking on a major pancreatic resection has previously been emphasized since ligation of a major hepatic arterial blood supply in a jaundiced patient may lead to fatal liver ischaemia.

Complications of angiography

Complications at the femoral puncture site include haemorrhage, haematoma, arterial occlusion, pseudoaneurysms and arteriovenous fistula. Embolization of thrombus formed at the puncture site, subintimal dissection of an atheromatous plaque with secondary thrombosis, distal embolization of catheter or guide wire fragments have all been described. Postangiographic renal failure is well documented, especially in patients with pre-existing renal disease. The injected contrast media are hypertonic and may be a real hazard, especially if the patient has been dehydrated for several hours. The jaundiced patient is particularly at risk from both bleeding problems and renal failure. They should be kept well hydrated and their coagulation abnormalities should be treated prior to angiography. The spectrum of adverse sequelae is indeed very broad and this invasive procedure should not be undertaken without good indication, by inexperienced personnel or where facilities are inadequate, but when planned and performed with care, abdominal angiography is an acceptably safe procedure. It is occasionally a useful part of the diagnostic armamentarium in patients with pancreatic disorders, but its value has dwindled and its place has been usurped by the recent improvements of other less invasive imaging techniques such as the modern CT scan, MRI and EUS.

Pancreatic angiography remains of value in two preoperative situations:

- arterial embolization of feeding vessels to large vascular tumours
- splenic artery embolization for lesions around the tail of the pancreas obstructing the splenic vein and causing splenomegaly and left-sided portal hypertension.

In both instances, the angiographic manoeuvre facilitates the subsequent surgical excision by significantly reducing intraoperative blood loss.

Endoscopic ultrasonography

EUS (Figure 27.11) has emerged as a very useful tool for visualizing the pancreas and peripancreatic structures by combining the advantages of endoscopy with ultrasonography. It can detect small tumours that are not visible on a CT scan and helps in assessing resectability of larger tumours by delineating their relationship to adjacent major vessels. It is however very dependent on the technical expertise and interpretation of an experienced individual. EUS, when available, has largely replaced diagnostic ERCP in the evaluation of the jaundiced patient. EUS-guided FNA of pancreatic masses and enlarged lymph nodes, and coeliac plexus nerve block are also being widely practised. EUS is used increasingly as the initial modality for evaluating pancreatic lesions. EUS alone is more sensitive (94%) than CT (69%) and MRI (83%) for detecting pancreatic lesions smaller than 3.0 cm. EUS has an accuracy of 93% in the prediction of local resectability compared with 60% for CT. The specificity of EUS is similar to angiography for the detection of vascular invasion.

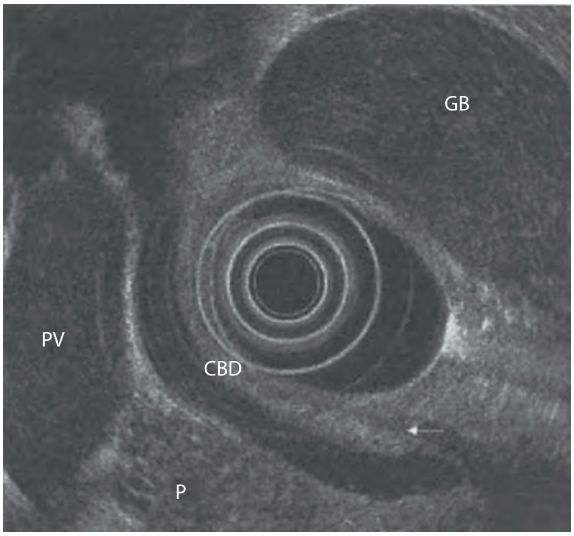


Figure 27.11 Endoscopic ultrasonography. CBD, common bile duct; GB, gallbladder; PV, portal vein; P, pancreas.

Ultrasound-guided FNA aspiration cytology is highly specific (99–100%) for the diagnosis of solid pancreatic lesions. The current diagnostic EUS-FNA criteria for pancreatic adenocarcinoma include increased cellularity; the predominance of one cell type; overlapping cells; multiple pleomorphic cells; tall cells with large nuclei (tombstones); and cells with increased nuclear/cytoplasmic and with coarse and clumped chromatin, macronucleoli and abnormal mitoses.

Increased cellularity is one of the important criteria in the differentiation of adenocarcinoma from chronic pancreatitis by EUS-FNA. However, the cellularity of a sample is influenced by several factors, including technique and the anatomic location of the lesion. Markedly cellular aspirates are obtained more frequently with EUS-FNA than with percutaneous FNA and for this reason the EUS-guided aspiration is preferred. Increased cellularity in FNA samples is also encountered in pancreatic islet cell tumours. Accurate diagnosis of an islet cell tumour and its differentiation from pancreatic adenocarcinoma, acinar cell carcinoma, and solid cystic pancreatic carcinoma is based on an assessment of morphological features and immunohistochemical staining.

Cytology alone is considered an insensitive test for the diagnosis of cystic pancreatic lesions; compared with solid pancreatic lesions, the overall sensitivity for FNA for establishing a definite diagnosis is reduced to 62% when applied to cystic pancreatic lesions. Nonetheless, EUS-FNA may be helpful but requires good technique for procurement and a very experienced pancreatic cytologist. Aspirates from microcystic adenomas yield hypocellular material containing cuboidal cells with bland nuclei and pale cytoplasm. Aspirates from mucinous cystic neoplasms are moderately cellular and demonstrate abundant mucinous material with glandular epithelial cells arranged in sheets and cohesive clusters. When present, the key cytological findings indicative of mucinous cystadenocarcinoma include moderate cellularity, loose clusters of cells, single cells, nucleoli, overt malignant nuclear features, and the presence of signet-ring cells. The features characteristic of intraduct papillary mucinous tumours (IPMTs) include large papillary groups with a fibrovascular core lying in pools of mucin. The neoplastic cells are columnar and show loss of cell polarity. The individual cells may demonstrate a wide range of morphological changes, from a preserved nuclear cytoplasmic (N/C) ratio and a regular nuclear membrane to marked anisocytosis, an increased N/C ratio and an irregular nuclear membrane with conspicuous nucleoli.

CEA, amylase, CA125 and CA19-9 from samples of the cyst contents have been used in the hope of increasing the sensitivity and specificity of the diagnosis. One recent study showed that non-invasive mucinous cystic neoplasms, irrespective of the degree of atypia, are positive for MUC5AC but negative for MUC1. In contrast, the lesions with an invasive component express MUC1. Thus the expression pattern of the MUC antigens may provide useful information for the determination of the invasive potential of a cystic mucinous lesion.

Pancreatic pseudocysts are distinguished from pancreatic cystic epithelial neoplasms by the predominance of histiocytes and inflammatory cells.

Magnetic resonance imaging

MRI makes use of the property of nuclear magnetic resonance to image nuclei of atoms inside the body. Modern MRI machines generate (1) powerful magnetic fields (1.5–3.0 T) to align the magnetized atoms of the body and (2) radiofrequency fields to alter the alignment of this magnetization. The response of the nuclei of the body tissues to this combined exposure is the production of a rotating magnetic field which is detected by the scanner and used to construct an image of the scanned region as strong magnetic field gradients cause nuclei at different locations to rotate at different speeds. Additionally, three-dimensional (3D) spatial imaging can be obtained by providing gradients in the x, y, z directions.

MRI provides good contrast between the different body tissues and is ideal for imaging the brain, muscles, heart and neoplasms. Its advantage over CT is that it does not expose the patient to ionizing radiation. There are various imaging techniques used in MRI but the most common are T_1 weighted and T_2 weighted imaging. T_1 weighted scans differentiate water (dark) from fat (bright) and use a gradient echo sequence with short echo time (T_E) and repetition (T_R). These techniques of MRI scanning can be used with or without a contrast agent (gadolinium) in which case T_1 scans are performed before and after administration of contrast. T_2 weighted MRI scans which use a gradient echo sequence T_E and long T_R are excellent for imaging fluid such as blood and bile etc. The special forms of MRI scanning include magnetic resonance angiography (MRA) based on the administration of paramagnetic contrast agents, magnetic resonance gated intracranial cerebrospinal fluid (CSF) dynamics (MR-GILD) used to investigated the CSF circulation in patients with obstructive lesions (e.g. hydrocephalus), functional MRI to study neural activity (mostly research) and MRCP. Increasingly MRI is also used for image-guided interventions and for guidance of tumour ablation by high-intensity focused ultrasound.

Technical advances, such as high field-strength (3.0 T) MRI with phased-array coils and ultrafast imaging have vastly improved the performance of MRI in the detection and staging of pancreatic cancer. MRI with a high gradient system and phased-array coils has a reported sensitivity of 95% for the diagnosis of pancreatic cancer, a positive predictive test of non-resectability 90% and a negative predictive test/verdict (tumour is operable) of 83%. MRI is also helpful in the detection or exclusion of pancreatic cancer in patients showing a prominent pancreatic head or uncinate process but no definite tumour on CT or ultrasound examination.

Magnetic resonance cholangiopancreatography

MRCP, which is based on T_2 weighted imaging, outlines the biliary tract with images which are similar to that obtained by ERCP by detection of the fluid contained in the biliopancreatic system (Figure 27.12).

Aside from its non-invasive nature, MRCP avoids the need for use of contrast and is therefore ideal for imaging of the entire biliopancreatic system in patients with known allergy to iodine-based contrast materials and in patients with a history of atopy. The diagnostic accuracy of MRCP is the same as that of ERCP for both benign and malignant disorders of the pancreas and biliary tract. However, ERCP has the advantage of permitting therapeutic interventions of the biliary and pancreatic ductal systems. In addition, ERCP enables inspection and biopsy of the periampullary region. 3D reconstruction from segmented MRI multislice imaging provides superb images of the biliopancreatic tracts.

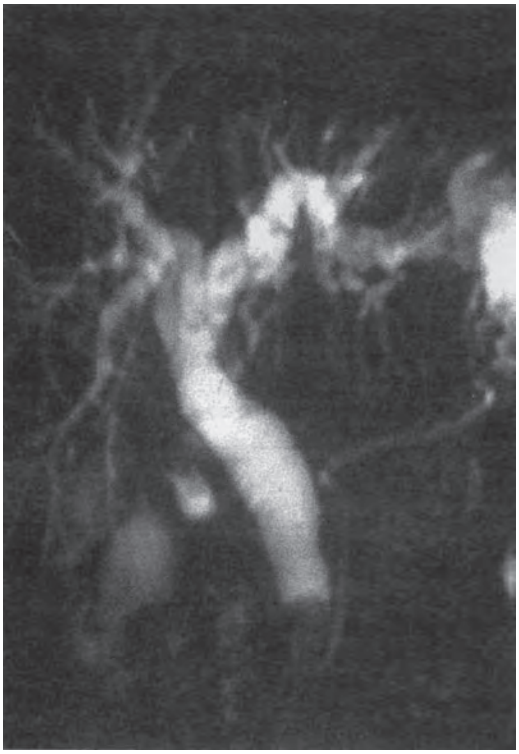


Figure 27.12 Magnetic resonance cholangiopancreatography showing an impacted stone in the distal common bile duct in a jaundiced patient. Note a normal duct of Wirsung.

Chemical analysis of the stool to demonstrate steatorrhoea

Stool examination for fat content is only useful as a screening test for malabsorption. Faecal fat excretion is not a valid measure of pancreatic dysfunction, since about 80% of pancreatic secretory capacity may be lost without any detectable change in the test. In clinical practice, pancreatic secretory deficiency states and malabsorption syndromes coexist in about 10–20% of patients. Even with more sophisticated tolerance tests, there is an overlap of the results obtained in malabsorption states and pancreatic insufficiency.

Direct measurement of pancreatic digestive and secretory capacity

Direct duodenal intubation and collection of pancreatic juice for analysis (duodenal drainage studies) following various stimuli is still widely practised largely as a research tool. The exact technique varies from institution to institution. In the Lund test, a meal of fat is given and the output of pancreatic lipase in the aspirated duodenal content is determined. The pancreatic secretory response to an injection of secretin (measurement of volume of juice and bicarbonate output) is often studied, but data interpretation is sometimes difficult. However, it is the most reliable test in detecting pancreatic exocrine insufficiency due to chronic pancreatitis or pancreatic carcinoma. Cytological examination of the duodenal aspirate during these maximal secretory tests sometimes documents the presence of malignancy in cases of pancreatic cancer.

Congenital anomalies of the pancreas

The pancreas may be totally absent but this condition is extremely rare and usually associated with other severe malformations that are incompatible with life. Pancreatic agenesis is associated with mutations of the *IPF1* gene on chromosome 13q12.1, which is a transcription factor essential for normal embryological development of the pancreas. The three congenital anomalies of the pancreas which are of surgical importance are ectopic pancreas, annular pancreas and pancreas divisum.

Ectopic (heterotopic, accessory, aberrant) pancreas

Ectopic pancreas (heterotopic, accessory or aberrant pancreas) is a rare developmental anomaly with varying estimates of 0.2–2% of the general population. It is defined as pancreatic tissue lacking anatomic and vascular continuity with the pancreas. These embryological rests are most commonly located in the distal stomach and duodenum followed by the jejunum, Meckel's diverticulum and less frequently the biliary tract, including the major duodenal papilla and other sites. Thus scattered pancreatic tissue has been reported in the gallbladder, colon, spleen, liver, bile ducts, mesentery and omentum. Some enterogenous cysts of the thorax have been reported to contain typical pancreatic tissue, including islets.

Ectopic pancreatic lesions are located most commonly in the submucosa (75%) and in the muscular layer or subserous coat in the remainder (Figure 27.13).

With the advent of widespread upper gastrointestinal endoscopy and improvements in contrast studies of the alimentary tract, ectopic pancreas of the stomach and duodenum is being more frequently recognized. The pathognomonic contrast radiological finding is a smooth, rounded filling defect with evidence of a tiny umbilication or even a small duct which may be outlined by a line of barium.

Ectopic pancreas is usually asymptomatic and discovered incidentally during endoscopy or laparotomy. However, it may become symptomatic because of the development of complications such as acute pancreatitis, pseudocyst, ulceration, gastrointestinal bleeding, gastric outlet obstruction, biliary obstruction with jaundice (ampullary lesions), intestinal obstruction and intussusception. In addition, there have been isolated reports of the development of malignant degeneration (pancreatic cancer) and islet tumours (duodenal insulinoma). When identified, the submucosal mass in the stomach or duodenum is initially misinterpreted as gastrointestinal stromal tumours as these account for 90% of gastric submucosal tumours or leiomyomas. There are no specific distinguishing features on CT scanning of lesions caused by ectopic pancreatic rests.

The symptoms are varied and include abdominal pain suggestive of ulcer dyspepsia, postprandial fullness and vomiting from gastric outlet obstruction, dyspepsia due to proven peptic ulceration, gastrointestinal bleeding and intestinal obstruction.

Annular pancreas

The exact embryological explanation for this malformation is debatable. The classic description of annular pancreas is almost

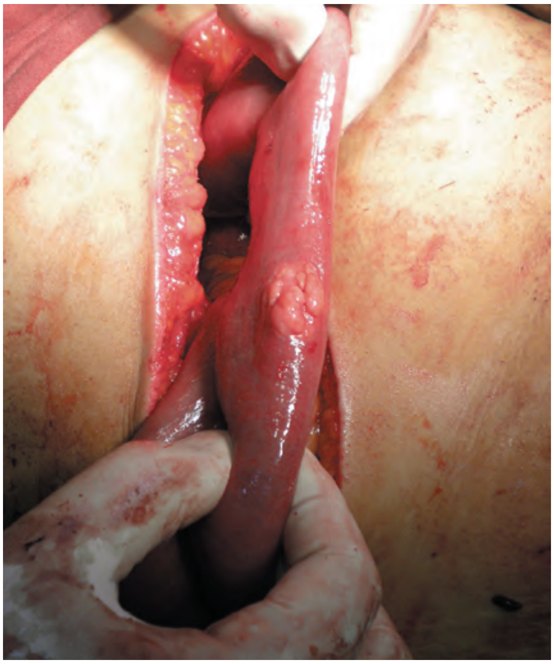


Figure 27.13 Ectopic serosal small bowel pancreas. (Courtesy of Professor Omar Shah, Srinagar, Kashmir, India.)

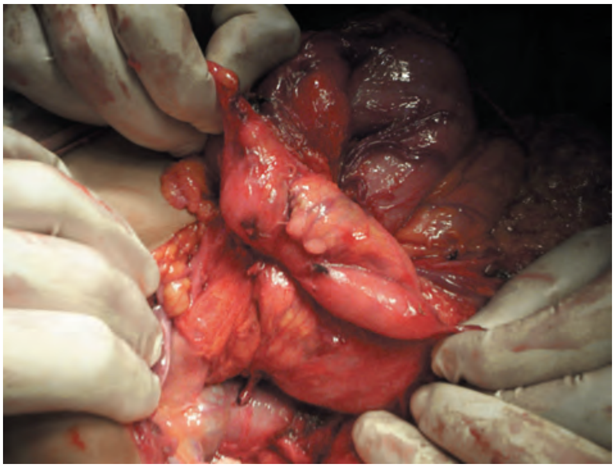


Figure 27.14 Complete annular pancreas. (Courtesy of Professor Omar Shah, Srinagar, Kashmir, India.)

invariably a ring of pancreatic tissue, continuous with the head of the pancreas, surrounding the second part of the duodenum proximal to the ampulla of Vater. However, the infra-ampullary location has also been documented. Eighty-five per cent of cases occur around the second portion of the duodenum, and the remaining 15% are scattered around the first or third part of the duodenum (Figure 27.14).

The pancreatic tissue is generally firmly attached to and embedded into the duodenal musculature; only rarely is it loosely attached and readily separable from the duodenum. There is usually a variable amount of hypertrophy of the proximal duodenal wall resulting from obstruction. Half of the reported cases have manifested themselves in the first year of life. Some believe that there is invariably an associated intrinsic atresia or stenosis of the duodenum. Other developmental anomalies are also present in about 60–70% of cases. They include Down syndrome, non-rotation and incomplete rotation of the mesentery, preduodenal portal vein, imperforate anus, oesophageal atresia with tracheoesophageal fistula and congenital heart disease. There is a high frequency of polyhydramnios in the mothers of those children who have significant duodenal obstruction at birth.

The exact incidence is unknown as many cases remain asymptomatic during life. However a postmortem study reported three cases in 20 000 autopsies. The condition is commoner in males. Vomiting is the main symptom of annular pancreas. It may begin as soon as the infant starts feeding or may appear several days later. Bile may or may not be present in the vomitus. Jaundice may be present and has been explained by back pressure of the distended duodenum on the common bile duct or involvement of the ampulla of Vater by oedema at the level of the stenosis. A plain film of the abdomen may show a distended stomach and a 'double-bubble' sign at the level of the duodenum. Contrast studies may be necessary to diagnose the condition.

About one-half of the cases reported present for the first time with symptoms between the ages of 21 and 70 years. The reason for such late manifestations of symptoms is generally attributed to inflammatory changes in the pancreatic ring. Duodenal ulcer, frequently reported with annular pancreas in the adult, has not been reported in infancy. The differential diagnosis in infancy rests between annular pancreas and other causes of duodenal obstruction, all of which require urgent operative relief, namely duodenal stenosis or atresia, compression of the duodenum by Ladd's bands, pyloric stenosis or volvulus in association with malrotation. An absolute differentiation between annular pancreas and duodenal atresia or stenosis is not possible without operation.

The operation of choice is either duodenoduodenostomy or duodenojejunostomy performed through the retrocolic route. Attempts at division of the pancreatic ring are attended with complications, including pancreatitis, pancreatic fistula, duodenal wall perforation and failures. Gastrojejunostomy is not a satisfactory operation since it inadequately decompresses the duodenum and has a high incidence of marginal ulceration. In patients with concomitant jaundice, an operative cholangiogram is mandatory before a decision about the need for and type of biliary diversion which can be made. Pancreas divisum and its implications are discussed in the section on pancreatitis.

Injuries to the pancreas

Pancreatic injuries occur infrequently in patients with abdominal trauma of all types and this may be accounted for by the relatively protected position of the gland in the retroperitoneum beneath the thoracic cage. Pancreatic trauma is often classified according to the source of the injury: (1) penetrating trauma; (2) blunt trauma; and (3) iatrogenic trauma. Although the pancreatic injury rarely, if ever, accounts for the early death of a patient, it adds significantly to the morbidity and late mortality, especially if it is recognized late.

Injuries to the pancreas are being encountered more frequently nowadays. This is attributed to the increasing incidence of motor vehicle accidents and civil violence. There is also some evidence that the compulsory use of seat belts in Western countries may have contributed to the increase in the incidence of blunt pancreatic trauma, since the gland may be ruptured by sudden compression against the lumbar spine. The overall mortality rate from pancreatic trauma is about 20%. Stab wounds have a mortality rate of about 8%, gunshot wounds 25% and shotgun wounds 60%. The mortality following blunt trauma from steering wheel injury is still approximately 50%. One-third of all pancreatic injuries are secondary to blunt trauma and two-thirds are due to penetrating trauma in the USA. The majority of early deaths result from massive haemorrhage and shock, from the associated injuries to major vascular structures.

Diagnosis

In less than 10% of patients with pancreatic trauma of all types is the pancreas found to be the only injured organ. This implies that over 90% of patients will have further internal injuries. In the clinical situation, patients with definite evidence of an injured intra-abdominal viscus and/or intra-abdominal haemorrhage following blunt or penetrating trauma to the abdomen are in need of an emergency laparotomy following adequate resuscitation. They should not be investigated further. The initial emphasis is placed on adequate resuscitation rather than an elaborate, time-consuming and often unrewarding investigation. The diagnosis and the extent of the injury is made intraoperatively.

On the other hand, patients with doubtful evidence of intra-abdominal injury after blunt trauma who are haemodynamically stable should be observed and carefully investigated as dictated by clinical circumstances. There are a small group of patients in whom an isolated pancreatic injury may easily be missed.

Classification of pancreatic injuries

Pancreatic injuries can be classified as follows:

- simple contusion
- parenchymatous laceration without major ductal injury
- major ductal disruption
- combined pancreatoduodenal injury.

Penetrating pancreatic trauma

Isolated pancreatic injury is very rare. Associated injuries include liver, stomach, major vascular structures, spleen, duodenum, colon and kidney. Early death is invariably due to the associated injuries which can lead to catastrophic, often uncontrollable, haemorrhage. Late death accounts for some 41% of all mortality and is attributable to the pancreatic injury, leading to intra-abdominal abscess, sepsis and multiorgan system failure.

The emphasis in the management of penetrating injuries of the abdomen is on resuscitation with minimal essential investigations. The diagnosis is made at laparotomy. A long midline laparotomy is performed to provide good access and exposure. Management of associated injuries takes precedence over any pancreatic injury. The underlying principles are (1) arrest haemorrhage, (2) control contamination from the gastrointestinal tract and (3) once the patient is stabilized, the lesser sac is widely opened and all peripancreatic haematomas are explored. If there is no major ductal disruption, selective pancreatic and peripancreatic debridement is performed only as necessary. Adequate external drainage is instituted and, if the patient develops a pancreatic fistula, it can be treated conservatively with a combination of total parenteral nutrition and somatostatin analogue.

Blunt pancreatic trauma

About 50% of patients with blunt pancreatic trauma will have an isolated pancreatic injury. If the patient is stable and the diagnosis of significant injury is doubtful, then the patient is observed and is investigated rationally. The only test that is of definite diagnostic value is CT with intravenous and oral contrast. Occasionally, a flat X-ray plate of the abdomen, EUS, ERCP and abdominal angiography may be useful. Serum amylase and lipase elevation or lack thereof are not usually helpful in this clinical situation, although if these enzymes are highly elevated, they will strengthen a diagnostic suspicion of pancreatic injury. If the patient is unstable and there is evidence of an injured abdominal viscus, the patient should be rapidly resuscitated and explored as for a penetrating pancreatic injury.

Treatment of major pancreatic ductal disruption

If the injury is well to the left of the mesenteric vessels, the simplest manoeuvre is to perform a distal pancreatectomy and splenectomy with oversewing of the cut end of the proximal pancreas. In children under 16 years of age, one may elect empirically, if conditions are favourable, to attempt preserving the spleen with its blood supply based on the short gastric vessels.

If the injury is close to or to the right of the superior mesenteric vessels, pancreatic preservation may be considered. The preservation of pancreatic tissue is often desirable, but should not become an irrational obsession. A partial pancreatic transection may be completed and the area debrided as necessary. The right side of the transection may be oversewn as before and the cut end of the left pancreas may be implanted into a Roux-en-Y loop of jejunum.

The decision to preserve pancreatic tissue and perform a pancreatoenteric anastomosis is based on (1) the magnitude of the associated injuries, (2) the general condition of the patient, (3) the degree of contamination and (4) the surgeon's expertise and experience in pancreatic surgery.

Pancreatoduodenal injuries

The duodenal injury has to be repaired on its own merits. Following duodenorrhaphy, the injury to the pancreatic head has to be evaluated. Minor lacerations with or without ductal injury can be appropriately debrided and drained. If the ductal injury and the trauma to the head parenchyma appear extensive, a duodenal diversion procedure can be performed. An incision is made in the distal antrum and the pylorus is closed with a running catgut suture. A distal gastrojejunostomy is performed using the antral incision to 'protect' the duodenal repair (Figures 27.15 and 27.16).

For more extensive trauma, the duodenal diverticulization procedure of Berne is sometimes advocated. This entails a distal gastrectomy, Billroth II gastrojejunostomy, T-tube insertion for drainage of the common bile duct, and duodenostomy tube drainage (Figure 27.17).

In all these extensive injuries, a duodenal and/or pancreatic fistula will eventually develop and this can be treated conservatively with total parenteral nutrition and a long-acting somatostatin analogue. If the fistula does not spontaneously close after a few months, then a secondary attempt to drain it into an isolated bowel loop can be considered.

Occasionally, especially with gunshot wounds to the right upper abdomen, extensive devitalization of the duodenum and pancreatic head is seen with avulsion of the common bile duct and uncontrollable haemorrhage. In such a situation, the surgeon may have no choice but to 'complete' the pancreatoduodenectomy if only to control haemorrhage, usually from the superior mesenteric vessels, the renal vessels and the inferior vena cava. It is not unusual in such situations to be forced to perform a concomitant right nephrectomy. An experienced surgical team is needed for this type of injury since the mortality is in excess of 30%.

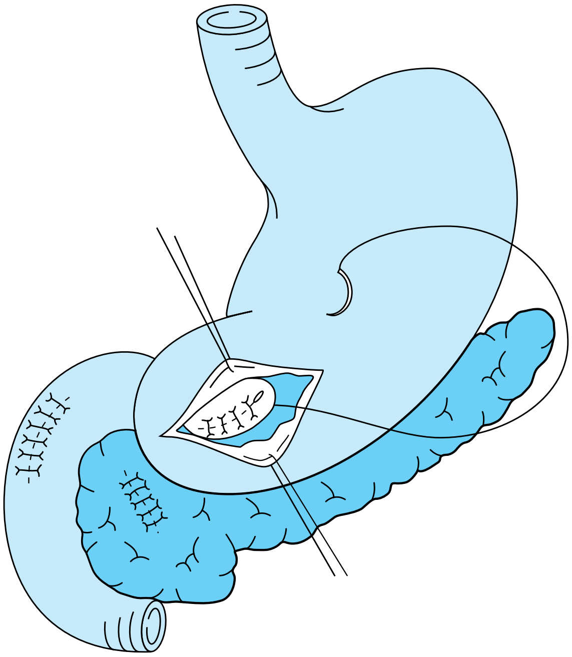


Figure 27.15 Pyloric exclusion procedure. The pylorus is oversewn through an antral incision.

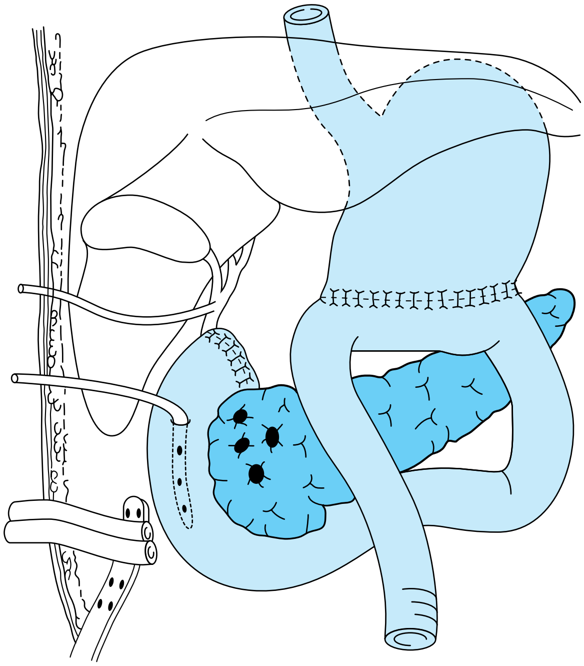


Figure 27.17 Duodenal diverticulization for major pancreatoduodenal injuries.

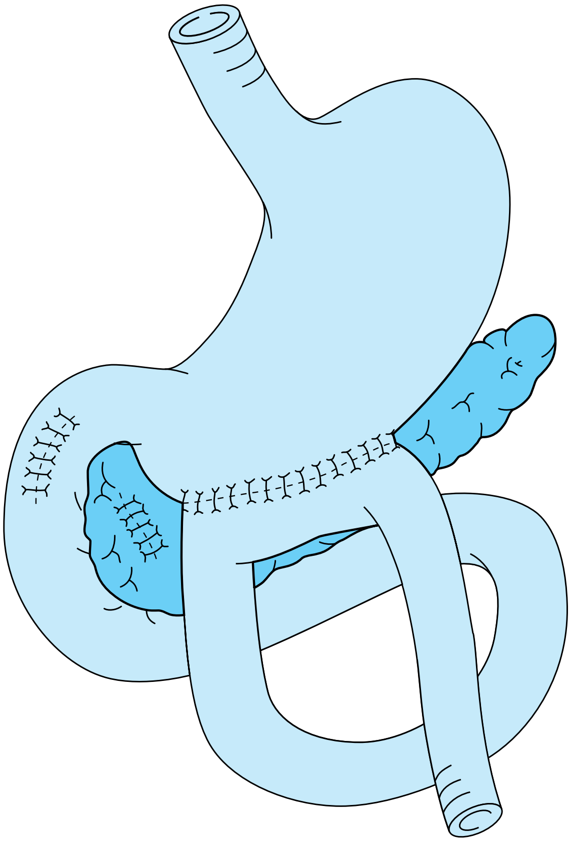


Figure 27.16 Pyloric exclusion procedure. A gastrojejunostomy is constructed through the previous antral incision.

Iatrogenic pancreatic trauma

The commonest cause of iatrogenic pancreatic trauma is ERCP with or without associated sphincterotomy and stent placement. Multiple pancreatic biopsies are another cause, whether performed percutaneously or intraoperatively. Occasionally, surgical operations on a large penetrating duodenal ulcer may also lead to severe pancreatic injury. Other surgical procedures which may lead to inadvertent pancreatic trauma are splenectomy, inexpert exploration of the common bile duct and transduodenal sphincteroplasty. All these invasive manoeuvres may lead to a fulminant pancreatitis with serious sequelae.

Pancreatitis

Definition and classification

The term ‘pancreatitis’ implies the presence of pancreatic inflammation and autodigestion. It can be classified according to its clinical presentation, according to the aetiological factors or according to the severity of the pathological process. Clinically, pancreatitis is referred to as acute or chronic. *Restitutio ad integrum* is the hallmark of the acute form, whereas persistence or progression of the disease or residual damage to the pancreas indicates the chronic variety. Nevertheless, if a patient recovers from acute necrotizing pancreatitis, normal recovery of pancreatic (endocrine and exocrine) function or morphology may never occur and the patient may be rightly labelled as having ‘progressed’ into chronic pancreatitis. Similarly, on occasion, recurrent attacks of acute pancreatitis due to any cause may merge into the chronic varieties. The emphasis on chronic pancreatitis is the absence of enzyme elevation and pancreatic exocrine and endocrine insufficiency. Patients with chronic relapsing pancreatitis, in addition to the same disorders of function, experience pain which, in the overt case, is constant with acute exacerbations against the background of pain. All types of acute or chronic pancreatitis may cause complications such as biliary tract obstruction, splenic vein thrombosis and duodenal obstruction.

Acute pancreatitis

Although, experimentally, acute pancreatitis has been produced in various animal models by promoting duodenopancreatic reflux, by the injection of bile salts, trypsin, etc., into the pancreatic duct, or by the occlusion of pancreatic arterial blood supply, the relevance of these models to the human clinical situation remains uncertain. Acute pancreatitis has been associated with a variety of clinical disorders, but the actual cause or mechanism which initiates the pancreatic autodigestion or which makes it either a self-limiting disease or a progressively fatal disease remains unclear. Acute pancreatitis developing in the course of viral infection, such as mumps and coxsackie B viruses, is usually self-limiting. So is the acute pancreatitis in association with mycoplasma pneumonia. Metabolic disorders, such as hypercalcaemia (usually associated with contraceptive pills), have been well documented. So has acute pancreatitis resulting from drug therapy, such as diuretics and steroids. Autopsy examinations of patients dying of such conditions as fulminant liver failure, hypovolaemic shock with renal tubular necrosis, transplantation and cardiac bypass surgery have revealed fulminant pancreatitis in a significant number of these patients. From the practical point of view, the two conditions most commonly associated with acute pancreatitis are alcoholism and biliary tract stone disease. The prognostic and therapeutic implications of acute pancreatitis due to these two associations are different. Gallstone pancreatitis usually occurs in patients older than 60 years, tends to be severe and is frequently accompanied by serious complications in the acute stage. However, recurrences may be prevented by early operative intervention or, if the patient recovers quickly, by appropriately timed elective surgical treatment of the biliary tract. By contrast, alcohol-associated pancreatitis tends to be recurring. Although the first attacks are usually quite severe, and can be fatal, subsequent attacks are commonly mild and carry low mortality. As time passes, the recurring attacks lead to progressive destruction of the gland and eventually the clinical picture merges into that of chronic pancreatitis or chronic relapsing pancreatitis.

Acute pancreatitis may be secondary to blunt or penetrating abdominal injury or may follow some invasive clinical test, such as ERCP and the now rarely performed translumbar aortography. Postoperative acute pancreatitis most commonly occurs after exploration of the common bile duct and is invariably due to injudicious, inexpert, forceful dilatation of the sphincter of Oddi or transduodenal sphincteroplasty. Distal pancreatitis may follow splenectomy.

Disease severity

The original Atlanta classification defined mild (oedematous) and severe categories of acute pancreatitis but is no longer appropriate in view of the improved understanding of the pathophysiology of acute pancreatitis and the importance of systemic complications on outcome and survival. It is now recognized that the severity of acute pancreatitis depends on several factors: (1) the number of organs that fail, (2) their onset and response to initial treatment and (3) the development of systemic (infectious) complications. Organ failure lasting for >48 hours is associated with increased mortality. The revised Atlanta classification includes all such patients as having severe disease as distinct from patients with transient organ failure who are considered to have mild/moderate acute pancreatitis. Local (peripancreatic) complications are not included in the category of severe acute pancreatitis unless accompanied by infection which is responsible for the high morbidity and mortality rather than the peripancreatic complications *per se*, as demonstrated by a retrospective study from the Mayo Clinic, which reported that patients with local pancreatic complications but without systemic complications had a negligible mortality (2%) despite an appreciable morbidity requiring a significant stay in intensive care and prolonged total hospital stay of 28 days. This study has been confirmed by a subsequent prospective study from the same institution and a prospective study from Spain. Thus, the revised Atlanta classification proposed by Petrov and Windsor includes a third category of moderate acute pancreatitis to include patients with local (peripancreatic) complications but no persistent systemic complications (previously graded as having severe acute pancreatitis). The Petrov–Windsor revised Atlanta classification recognizes a fourth category, which includes patients with both local peripancreatic and systemic complications (organ failure) during the course of acute pancreatitis. This category of extremely severe or critical acute pancreatitis identifies patients at high risk of mortality from the illness (28–58%) in various reports.

The classification and definitions of the four categories of the severity of acute pancreatitis are shown in Table 27.1.

Clinical features

The clinical manifestations of acute pancreatitis are protean since it can mimic any other abdominal emergency. In fact, it can coexist with some such as acute cholecystitis. It can occur at any age, but is rare in children and in very young adults. In the latter group, it is usually associated with infections,

Table 27.1 The classification and definitions of the four categories of the severity of acute pancreatitis

Severity category	Local complications		Systemic complications
Mild	No peripancreatic complications	and	No organ failure
Moderate*	Sterile peripancreatic complication	or	Transient organ failure
Severe*	Infectious peripancreatic complication	or	Persistent organ failure
Critical	Infectious peripancreatic complication	and	Persistent organ failure

*Severity is graded on the basis of more severe local or systemic complication (e.g. sterile pancreatic necrosis without organ failure has to be graded as 'moderate'; sterile pancreatic necrosis with persistent organ failure has to be graded as 'severe').

trauma, parasites or drugs or may be hereditary. Alcohol-related pancreatitis is usually found in the young adult less than 40 years of age, whereas the form associated with biliary tract disease manifests itself mainly in middle-aged and older persons.

The onset of symptoms may follow excess intake of food or alcohol but this is not invariable. Pain is the major initial symptom in over 90% of patients. It may vary in degree from mild to severe and is characteristically sudden in onset persisting for 12–48 hours or more. The wide spectrum of the intensity of the pain should be emphasized. Like many patients with pancreatic pain, there is a tendency for the patient to bend forward and assume various postures in order to obtain relief. Occasionally, even severe disease may be totally painless. The location of the pain is usually in the mid-epigastrium but can occur or radiate anywhere and may be diffuse or confined mainly to the back. Other common symptoms are nausea, vomiting, retching and hiccups. Less frequently, diarrhoea, dyspnoea, cyanosis, haematemesis and melaena may appear early. The bleeding may be due to erosive ulcers, bleeding diathesis, left-sided portal hypertension or erosion of the inflammatory process into the gut.

The physical findings may be minimal, in striking contrast to the severity of the symptoms. Obtundation, fever, tachycardia, epigastric tenderness and muscle guarding are frequent. Shock may be the initial symptom and in critical acute pancreatitis may be profound. Mild jaundice may be observed. Abdominal distension is often present in the early phase of the disease and is the result of a diffuse or localized paralytic ileus. Several days later, a palpable epigastric mass may be felt, indicating the development of peripancreatic complications, e.g. pseudocyst or marked peripancreatic fat necrosis. Other late signs (rarely encountered) include bluish discoloration of the skin around the periumbilical area (Cullen's sign) or in the loins (Grey Turner's sign); these indicate ecchymosis caused by seepage of blood arising from acute necrotizing pancreatitis along fascial planes. Another rare skin manifestation is nodular fat necrosis simulating erythema nodosum or Weber–Christian disease. Rarely, polyarthritis or bone pain may be observed and is also attributable to fat necrosis. Thrombophlebitis of the leg veins is only rarely seen. Abnormal physical findings on chest examination are found in less than one-third of cases, but a plain chest radiograph may reveal abnormalities in a greater proportion of patients. These include elevation of the diaphragm, basal atelectasis or pneumonitis, left-sided or bilateral pleural effusion.

Diagnostic procedures

The haematocrit is high in most patients at the onset of the disease, reflecting fluid loss related to the inflammatory process and vomiting. Leucocytosis with a white blood count of $10–20 \times 10^9$/dL is almost invariably present.

Serum enzyme elevations

The key to the diagnosis of acute pancreatitis remains an elevation of pancreatic enzymes in the circulating blood or urine. Although hyperamylasaemia is a non-specific finding, the level of the serum amylase in various extrapancreatic conditions rarely attains a value five times above normal and this is commonly seen in primary acute pancreatitis. Normal serum amylase levels may be obtained if the determination is carried out 3 days or more after the onset, if previous attacks have completely destroyed the glandular tissue or if the current attack is associated with massive destruction of the gland. Persistence of hyperamylasaemia for 10 days or longer may indicate a continuation of the acute inflammation or the development of local peripancreatic complications and may be accompanied by pancreatic ascites.

In view of the problems involved in the interpretation of hyperamylasaemia, measurement of isoenzymes of amylase and determinations of the renal amylase creatinine clearance ratio are used. The P-type isoamylase cannot be detected after extirpation of the pancreas and hence is only of pancreatic origin. Its measurement is unnecessary and time consuming for routine purposes but it is useful in establishing the cause of a persistent hyperamylasaemia. In acute pancreatitis, the relative renal clearance of amylase increases compared with that of creatinine. It has been suggested that a clearance of amylase over the clearance of creatinine of greater than four is found in the first few days after an attack of acute pancreatitis and that this elevated ratio persists longer than abnormalities of blood or urinary enzyme levels. The view that the amylase–creatinine clearance is specific for acute pancreatitis has now been shown to be untrue since an elevated clearance ratio is found in many other conditions, such as diabetic ketoacidosis, burns and chronic renal failure, whereas many patients with acute pancreatitis may have normal clearance ratios.

Estimation of lipolytic activity in the blood, urine, pleural and peritoneal effusions has been advocated on the grounds that lipase is more specific for pancreatic disease than amylase and its level remains elevated for longer periods. The recent introduction of radioimmunoassay for trypsin has resulted in reports demonstrating high levels of immunoreactive trypsin in the circulation and in serous effusions, in patients with acute pancreatitis. The diagnostic or prognostic value of immunoreactive trypsin assays compared with other enzymes has yet to be confirmed. Immunoreactive elastase levels have also been reported and are said to be elevated in patients with acute pancreatitis showing good correlation with levels of amylase.

Urinary amylase and lipase

Amylase clearance is increased about threefold for 1–2 weeks in patients with acute pancreatitis whose renal function is unimpaired. Thus, increased urinary output of amylase may persist for several days after normalization of the serum amylase levels. Determination of the ratio of amylase–creatinine clearance permits the diagnosis of acute pancreatitis even in the presence of renal insufficiency. Timed urinary amylase output (usually expressed in terms of total activity per hour) and amylase clearance studies are particularly of diagnostic value after serum amylase has returned to normal. Urinary lipase determination has been abandoned since it is debatable whether lipolytic activity is detectable in the urine.

Hyperglycaemia and glycosuria

Transient hyperglycaemia has been observed in a varying proportion of patients with acute pancreatitis. One of the mechanisms may be the release of glucagon from the damaged pancreatic cells. Usually the hyperglycaemia disappears with remission of the acute pancreatitis but, occasionally, following extensive parenchymal damage, diabetes mellitus may ensue.

Hypocalcaemia

The incidence, mechanism and prognostic implications of hypocalcaemia (serum calcium below 1.9 mmol/L or 7.5 mg/100 mL) are still controversial. Some believe that hypoalbuminaemia accounts for most of the measured hypocalcaemia. Deposition of calcium in areas of fat necrosis, release of glucagon with secondary hypercalcitonaemia and hypomagnesaemia have all been suggested as pathogenetic factors. Impairment of parathyroid gland function has also been reported. Whatever its mechanism, hypocalcaemia complicating acute pancreatitis is occasionally manifested by tetany and usually denotes the presence of the severe form of the disease.

Methaemalbuminaemia

The presence of this haemoglobin derivative in the serum suggests severe haemorrhagic pancreatitis. Elevated methaemalbumin levels, however, are not specific for necrotizing pancreatitis, because they can occur with other necrotizing intra-abdominal processes.

Blood coagulation tests

A rise in circulating trypsin may result in an increased antithrombin activity, but this is not a reliable test for acute pancreatitis. The persistence of elevated serum fibrinogen levels, especially after the second week, is said to indicate either a severe form of acute pancreatitis or the onset of complications.

Hyperlipidaemia

Hyperlipidaemia, especially hypertriglyceridaemia, occurs in about 5–10% of patients with acute pancreatitis. The lipid rise may be primary and/or the cause of the acute pancreatitis or it may be secondary to the pancreatitis. Long-term studies of blood lipids after subsidence of acute pancreatitis may be necessary for the differentiation of primary and secondary hyperlipidaemia. It has been suggested that many of the patients with so-called ‘secondary hyperlipidaemia’ may have a primary disorder of lipid metabolism which is unmasked by the attack of pancreatitis.

Electrocardiogram

Electrocardiogram (ECG) changes simulating myocardial infarction have been reported in patients with acute pancreatitis. The commonest ECG changes are S–T segment elevation or depression, inversion of T-waves, and extended T-wave negativity. These changes disappear rather quickly unless there is coexistent myocardial ischaemia or infarction or there are other cardiac complications of acute pancreatitis, such as pericarditis. The exact mechanisms underlying the ECG changes include several factors, such as myocardial damage due to shock, electrolyte disturbance, excessive parenteral fluid replacement, effect of severe pain on coronary artery disease, influence of circulating pancreatic trypsin, and vagally mediated reflexes from the pancreas to the heart.

Plain abdominal and chest radiography

A plain radiograph of the abdomen and chest may show evidence of pneumoperitoneum, thus excluding the diagnosis of acute pancreatitis and indicating the need for emergency operation for a perforated viscus. It may show abnormalities suggestive of acute pancreatitis in about 50% of cases. The radiological signs include intestinal distension in the region of the pancreas (sentinel jejunal loop, colon cut-off, duodenal ileus) or a generalized paralytic ileus. Haziness in the flat plate of the abdomen is caused by retroperitoneal fluid accumulation and may be associated with obliteration of the psoas outline. Other signs are elevation of the left diaphragm caused by basal atelectasis or subdiaphragmatic fluid collection, and pleural effusions.

Imaging techniques used in patients with acute pancreatitis

Contrast-enhanced CT is the standard imaging technique for detection of acute pancreatitis. It has 78% sensitivity and 86% specificity for severe acute pancreatitis. CT generally is not indicated for patients with mild uncomplicated disease but is reserved for patients with severe disease or worsening of the clinical and biochemical profile. Costs aside, the potential risk of contrast medium morbidity is the reason for limiting its use only to severe or deteriorating disease.

ERCP is helpful in evaluating less common causes of pancreatitis, e.g. microlithiasis, sphincter of Oddi dysfunction, pancreas divisum, etc. Urgent ERCP is indicated in patients at risk or with evidence of (1) biliary sepsis, (2) severe pancreatitis with biliary obstruction, (3) cholangitis, (4) rising bilirubin with worsening/persistent jaundice and (5) worsening pain in patients with an abnormal ultrasound examination. These patients need urgent biliary decompression. Several clinical trials have confirmed that, in patients with severe gallstone pancreatitis, morbidity and mortality are reduced with early interventional ERCP (papillotomy, stenting, etc.).

Although MRI is not commonly used, it is certainly indicated when improved visualization of peripancreatic inflammation, necrosis or fluid collections is needed.

MRCP is used preoperatively to determine which patients would benefit from ERCP as it has high (81–100%) sensitivity for detecting ductal calculi. Moreover, recent data indicate it to be as accurate as contrast-enhanced CT in predicting the severity of pancreatitis and identifying pancreatic necrosis (83% sensitivity and 91% specificity) in severe disease. EUS is useful in obese patients and in patients with ileus. It can determine which patients would benefit from interventional ERCP and can assist with endoscopic transmural cyst and abscess drainage.

Differential diagnosis

The clinical picture of acute pancreatitis tends to change with evolution of the disease. Early (2–3 hours) after the onset, symptoms are often suggestive of acute cholecystitis and, indeed, acute cholecystitis may coexist with acute pancreatitis. After 6–8 hours a perforated duodenal ulcer or acute appendicitis may be simulated. After 2–3 days, the clinical features may mimic those of an intestinal obstruction because there is marked abdominal distension and ileus. If the patient presents with profound cardiovascular collapse, myocardial infarction, acute aortic distension, ruptured aortic aneurysm, mesenteric infarction or even massive pulmonary embolism may be initially suspected. Most of these conditions may be associated with elevation of serum amylase and often of lipase. Improved diagnostic ability will result if the clinical limitations in assessment are fully appreciated. Thus, the diagnosis of acute pancreatitis is based on 'thinking about it as a possibility' and excluding other sources of abdominal pain.

Predictors of severity of acute pancreatitis

It is important to identify patients with severe to critical acute pancreatitis early (day 1 or 2 of admission) because timely administration of broad-spectrum, pancreas-penetrating antibiotics decreases the incidence of pancreatic and peripancreatic infection. Two well-controlled studies have shown a convincing benefit of imipenem when administered within 48 hours of onset and maintained for 10 days to 2 weeks in patients with severe pancreatitis. The incidence of pancreatic/peripancreatic infection is halved, from 40% to 20%. Predictors of severity fall into three main categories: clinical (long-established), biochemical and radiological (based on CT).

Clinical scoring systems

The sensitivity of clinical judgement, even in experienced hands, is low (about 40%) for predicting the severity of an acute attack of pancreatitis. Because of the subjective nature of the clinical acumen, Ranson developed 11 objective criteria and postulated that the risk of death and/or developing major complications may be estimated objectively by five parameters on admission to hospital and six parameters within the initial 48 hours of admission.

The clinical predictors of severity of acute pancreatitis according to Ranson are shown in Table 27.2.

Table 27.2 Clinical predictors of the severity of acute pancreatitis, according to Ranson

	Ranson criteria	
	Non-biliary	**Biliary**
Admission		
Age (years)	>55	>70
WBC	>16 000	>18 000
Glucose	>200	>220
LDH*	>350	>400
AST*	>250	>250
First 48 hours		
↓Hct	>10	>10
↑Urea	>5	>2
Ca^{2+}	<8	<8
PaO_2	<60	<60
Base deficit	>4	>5
Fluid requirement	>6	>4

*International units/mL.
AST, aspartate transaminase; Hct, haematocrit; LDH, lactate dehydrogenase WBC, white blood cell count.

In patients with fewer than three of these 11 prognostic factors the mortality rate is low (0.9%), with three or four factors, 18%; with five or six factors, 50%; and with more than six factors, 90%.

Imrie proposed the Glasgow (Imrie) criteria which are based on eight of Ranson's factors. The presence of any three of the eight signs at anytime within 48 hours of admission indicates a severe form of the disease (Table 27.3).

While these combined criteria are better than clinical judgement at predicting the severity and prognosis of acute pancreatitis, they pose some limitations: (1) they require data from the initial presentation and 48 hours to mature; (2) evaluation of the patient after 48 hours may be problematic; and (3) the criteria are not pancreas specific. They measure the systemic response to an inflammatory process in and around the pancreas and do not predict specific complications, e.g. severe necrosis, infection, pseudocyst and organ failure.

Of the clinical grading systems, the APACHE II system is currently the best and easiest scoring system available to assess the severity of acute pancreatitis and can be used *at any time* in the course of the disease. It is more sensitive and specific than the Ranson or Imrie criteria and is the clinical scoring system of choice. The APACHE II system allocates three sets of points: A, B and C. The A points assess clinical parameters, e.g. vital signs, electrolytes and arterial blood gases. The B points are allocated according to chronological age. The C points take into consideration comorbidity or chronic health of the patient. The APACHE II score is the sum of the A, B and C points. If it exceeds 9, the patient has severe acute pancreatitis. If the score increases after admission, the mortality is extremely high (Table 27.4).

Biochemical markers

Several biochemical 'markers' of the severity of acute pancreatitis have been studied (Box 27.1).

Ideally, any serum or urinary marker must be pancreas specific, originating from the organ producing the marker. Serum amylase or lipase is of *no* value. Thus far, four pancreas-specific markers appear promising in discriminating between mild and severe pancreatitis: (1) trypsin-activated peptide is a five amino acid peptide released by activation of trypsinogen and can be measured in the urine; early reports indicate a sensitivity of 80% and specificity of 90%; (2) phospholipase A_2 tends *not* to decrease in severe acute pancreatitis but the methodology of the assay has

Table 27.3 Clinical predictors of the severity of acute pancreatitis, according to Imrie

	Glasgow (Imrie) criteria
Age (years)	>55
WBC	>15 000
Glucose	>180
LDH*	>600
Urea	>96
Ca^{2+}	<8
PaO_2	<60
Albumin	<3.2

*International units/mL.
LDH, lactate dehydrogenase; WBC, white blood cell count.

Table 27.4 APACHE II scoring system

ColumnA		Column B		Column C	
Parameter	**Points**	**Age**	**Points**	**Chronic health**	**Points**
Temperature	0–4	<44	0	Cirrhosis	2 or 5
Blood pressure	0–4	45–54	2	Immunosuppression	2 or 5
Heart rate	0–4	55–64	3	NY Heart Association class IV	2 or 5
Respiratory rate	0–4	65–74	5	COPD	2 or 5
PaO_2	0–4	≥75	6	Renal dialysis	2 or 5
Arterial pH	0–4				
Na^+	0–4				
K^+	0–4				
Cr	0–4				
Hct	0–4				
WBC	0–4				
HCO_3^-	0–4				

Glasgow coma score

APACHE II score: A points + B points + C points >9 points suggests severe pancreatitis.

COPD, chronic obstructive pulmonary disease; Cr, creatinine; Hct, haematocrit; WBC, white blood cell count.

BOX 27.1 Biochemical markers of severity of acute pancreatitis

Pancreas specific

- Trypsin-activated peptide
- Phospholipase A_2
- Pancreatitis-associated protein
- Trypsinogen-2
- Ribonuclease
- Elastase 1
- Antichymotrypsin
- Carboxylic ester hydrolase
- Procarboxypeptidase β

Non-pancreas specific

- C-reactive peptide
- Procalcitonin

PMN elastase

- α_1-Antitrypsin
- α_2-Macroglobulin
- Methaemalbumin

been questioned and its value is uncertain; (3) the pancreatitis-associated protein is a protein purified from the pancreatic juice of rats with experimental pancreatitis – the human counterpart correlates well with the severity of pancreatitis, but the changes in concentration take days to a week to become established; (4) procalcitonin, which appears to be an early marker of severe disease; (5) serum trypsinogen-2, which is reported to be 91% sensitive and 71% specific as an index of severe disease, but further confirmation is needed. At present, none of these pancreas-specific markers can be considered to be of value in routine clinical practice.

The most extensively studied non-specific marker of acute pancreatitis is the C-reactive protein (CRP). This acute phase protein can reliably differentiate mild from severe acute pancreatitis but it is only useful after the second to third day of the disease. Its sensitivity is high (about 95%) but its specificity is low (about 50%). CRP concentration in excess of 100 mg/dL usually indicates necrotizing pancreatitis. However, its clinical value is diminished because it requires a few days to help in the identification of severe disease.

Another approach to staging acute pancreatitis is to monitor interleukins (ILs) or complement activation as a monitor of the local and systemic inflammatory response (Box 27.2). Tumour necrosis factor, IL-2, IL-6 and IL-8 have been extensively studied in experimental pancreatitis. Clinically, IL-6 and IL-8 seem to be as useful as CRP levels and peak about 1 day earlier. However, their value in routine clinical practice is as yet uncertain. Clinical studies with complement activation factors C3 and C4 have thus far been disappointing.

Radiological assessment: CT Severity Index

This is the sum of the CT grade of the condition of the pancreas (0–4) and the necrosis score (0–6).

- CT grade
  - A is normal pancreas = 0
  - B is oedematous pancreas = 1
  - C is B plus mild extrapancreatic changes = 2
  - D is severe extrapancreatic changes plus one fluid collection = 3
  - E is multiple or extensive fluid collections = 4
- Necrosis score
  - None = 0
  - >One-third = 2
  - <One-third but less than one-half = 4
  - >One-half = 6

Some studies have indicated superiority of the CT Severity Index in predicting the severity of acute pancreatitis compared with the other clinical systems (Ranson's criteria and APACHE II). A CT Severity Index score of 5 or greater correlates with prolonged hospitalization and higher mortality and morbidity rates. However caution is necessary with the interpretation of these comparative studies as, in most cases, the CT Severity Index was conducted 72 hours after admission, whereas the APACHE II scale and Ranson's criteria scores were calculated earlier – at 24 and 48 hours, respectively, although one study in which the CT Severity Index scores were obtained within 48 hours of admission reported similar findings. Thus the CT Severity Index is undoubtedly useful if available, but is unlikely to replace

BOX 27.2 Interleukins as indicators of the local and systemic inflammatory response to acute pancreatitis

- Tumour necrosis factor
- Interleukin 2
- Interleukin 6*
- Interleukin 8*
- Serum complement factor C3
- Serum complement factor C4

*Markers of macrophage activation.

clinical scoring systems in the short term, especially APACHE II, which demands constant monitoring of the patient, an integral component of medical treatment. Scoring systems apart, signs of organ failure within 24 hours of admission is a bad prognostic sign as it is invariably associated with increased mortality.

Treatment of acute pancreatitis

Medical management

The treatment of acute pancreatitis is essentially medical (conservative) in the first instance and is based on the following principles:

- volume replacement as required
- control of pain
- close monitoring of haemodynamic status, organ failure and development of peripancreatic complications
- maintenance of nutritional state
- active intervention (surgical, endoscopic, radiological) for infective complications.

There is now good evidence that total enteral nutrition is superior to parenteral nutrition in these patients, although some debate still exists on the type of total enteral nutrition (oral, nasogastric or nasojejunal) that is most effective in improving patient outcome. Nonetheless randomized studies have confirmed that enteral feeding reduces morbidity and mortality, possibly by prevention of bacterial translocation from the gut and thus infection of any necrosis. However, enteral nutrition cannot be used in patients with paralytic possible, when nutritional status is maintained by parenteral feeding. The issue of prophylactic antibiotic therapy initiated on admission remains debatable. One recent randomized trial (ciprofloxacin, metronidazole) against placebo reported no significant difference in the incidence of infected pancreatic necrosis (IPN), systemic complications and mortality. This is in disagreement with earlier trials and a recent systemic review which reached the opposite conclusion – prophylactic antibiotic therapy reduces infectious complications (pancreatic and extrapancreatic) and mortality. As prophylactic antibiotic has its risks (promotion of fungal infections and drug resistance), the decision on its use is difficult in view of the existing conflicting data. Selective use in patients with severe disease or bad prognostic criteria would seem to be a reasonable compromise.

The mainstay of the medical treatment of acute pancreatitis entails correction of hypovolaemia by replacement of fluid, electrolytes, blood or plasma. The continuing activity of the pancreatitis and/or the development of complications may be assessed by serial clinical examination, monitoring of serum and/or urinary amylase, and by regular CT scan of the pancreas. Current evidence suggests that anticholinergics, glucagon or Trasylol, and octreotide have no place in the management of patients with acute pancreatitis.

Medical management also includes close monitoring of vital signs, hourly urine output and central venous pressure measurements. Arterial blood gases must be frequently measured during the first few days as clinically occult respiratory failure is common. Pulmonary arterial pressure monitoring is valuable in patients with large volume requirements and especially in those with associated cardiopulmonary disease or respiratory failure. Such careful and repeated evaluation will also permit accurate differentiation between acute pancreatitis and other acute illnesses in the majority of cases. Selected patients may also need additional investigations, such as contrast studies of the alimentary tract, isotope biliary scanning to exclude with absolute certainty gastrointestinal perforation or obstruction, acute biliary tract disease, or mesenteric infarction.

Surgical management

Early surgical intervention

There are three situations when early operative intervention is indicated in a patient with suspected acute pancreatitis.

- When the diagnosis is in doubt and a perforated or a gangrenous viscus cannot be excluded.
- Patients with known biliary stone disease who develop an attack of acute pancreatitis which does not improve within 48–72 hours, especially if there is evidence of biliary obstruction and/or cholangitis. In this clinical setting, one must suspect an impacted stone at the ampulla and emergency ERCP with endoscopic papillotomy and stone extraction is indicated. If this fails or if an experienced endoscopist is not available, then early operation to relieve ampullary obstruction is mandatory.
- The patient who recovers following conservative management of acute gallstone pancreatitis has a high risk of recurrence. Early operative treatment of the cholelithiasis should be undertaken as soon as the pancreatitis has subsided, preferably during the same hospitalization.

It should be stressed that early abdominal exploration with pancreatic drainage or necrosectomy has an increased mortality in patients with severe pancreatitis. Peritoneal lavage by catheters placed percutaneously, coupled with adequate resuscitation, is effective in the management of early cardiovascular, respiratory and renal complications of severe pancreatitis, reducing early deaths from those causes. However, improvement in overall mortality following peritoneal lavage has been disappointing since lavage does not reduce the late deaths from pancreatic and peripancreatic necrosis and sepsis. It should be emphasized that peritoneal lavage and laparotomy are emergency resuscitative measures for the patient who is in refractory cardiovascular collapse and respiratory embarrassment. They should not be used indiscriminately for all cases of acute pancreatitis. If laparotomy is performed, the addition of a gastrostomy and jejunostomy confers several advantages. The former always allows gastric decompression and drainage which may be needed for several weeks without the discomfort of a nasogastric tube. The latter provides a route for enteral nutrition and obviates the use of intravenous alimentation which has the potential risk of catheter sepsis. It was rightly pointed out that the two surgical pitfalls in acute pancreatitis are (1) to operate too early and do too much and (2) to operate too late and do too little.

Any patient who has persistence or reappearance of inflammatory manifestations of acute pancreatitis must be suspected of developing a pancreatic pseudocyst or a pancreatic abscess. Clinically, fever, pain, tenderness, a palpable mass, leucocytosis and hyperamylasaemia are inconstant features. A positive blood culture is not usually obtained in the early development of a pancreatic abscess. If antibiotics are used, the fever may be partly masked. A plain radiograph of the abdomen may demonstrate the mottled 'soap bubble' appearance suggestive of a retroperitoneal abscess in less than 20% of cases. CT is an invaluable tool in these cases and obviates the need for contrast studies of the gastrointestinal tract.

Apart from the eradication of biliary tract disease following an attack of acute pancreatitis which has subsided, the role of operative intervention late in the course of acute pancreatitis is for the treatment of complications:

- IPN/abscess formation
- pseudocyst formation
- haemorrhage resulting from pseudoaneurysms or sectorial (left-sided) portal hypertension.

Management of pancreatic pseudocyst

A pancreatic pseudocyst is a collection of fluid in the lesser sac and its walls have no recognizable epithelial lining. However, at some point, most pseudocysts connect with the pancreatic glandular tissue or ductal system and the discharge of fluid into the cyst is maintained via the connection. Microscopic examination of this area of a pseudocyst will reveal some epithelial lining. The majority of pancreatic epithelial cysts are neoplastic (Table 27.5)

The differentiation between a true cyst and a pseudocyst is only possible by histology. However, the epithelial lining of true cysts may atrophy due to overdistension or infection. Hence, the absence of histologically recognizable epithelium in the cyst wall on biopsy may be of no significance. Two types of pancreatic pseudocysts can be differentiated. Acute pseudocysts (acute peripancreatic effusions) follow an established acute attack of pancreatitis. They should be managed expectantly for 4–6 weeks. Spontaneous resolution often occurs and surgical therapy is more satisfactory if the cyst wall is allowed to mature. *Chronic pseudocysts* are usually asymptomatic and, often, no recent attacks of acute pancreatitis can be identified. Spontaneous resolution is rare and delay only invites the high risk of complications. The commonest cause of pancreatic pseudocysts in children is blunt abdominal trauma. Pancreatic pseudocyst is a common complication of chronic pancreatitis occurring in 20–40% of cases.

Table 27.5 Cysts of the pancreas

True cysts	Pseudocysts
Benign cysts	
Enterogenous cyst	(1) Acute pseudocyst following acute pancreatitis
Retention (solitary) cyst	(2) Chronic pseudocyst following acute pancreatitis
Dermoid cyst	(3) Chronic pseudocyst associated with chronic pancreatitis
Lymphoepithelial cyst	(4) Miscellaneous peripancreatic cysts
Cystic fibrosis	– hydatid cyst
von Hippel–Lindau disease	– choledochal cyst
Endometrial cyst	– postpancreatitis necrotic collections
Neoplastic cysts	
Serous cystadenoma	
Mucinous cystadenoma	
Cystadenocarcinoma	
Papillary cystic neoplasm	
Solid malignancies with central necrosis and cavitation	

The other parameters which are helpful in the management of pancreatic pseudocysts are listed below.

- Size – pseudocysts less than 5 cm in diameter may be observed and expected to resolve in many instances. Those greater than 7.5 cm in diameter will probably need surgical drainage.
- The development of symptoms is indicative of an impending complication such as rupture, haemorrhage and infection.
- Maturity – acute pseudocysts should be allowed to mature for 4–6 weeks to allow the cyst wall to mature since this facilitates internal drainage.
- Vascular complications – visceral angiography delineates a subgroup of patients with vascular complications associated with acute pancreatitis. These include pseudoaneurysms and left-sided portal hypertension from splenic vein thrombosis. If arterial complications are suspected, an arteriogram is helpful especially if preoperative arterial embolization is contemplated. The presence of portal hypertension is an indication for splenectomy.
- The site of the pseudocyst is another important parameter which helps the surgeon decide on the method of internal drainage. Retrogastric cysts which are enlarging anteriorly are best treated by a posterior cystogastrostomy. Cysts around the head of the pancreas close to the duodenum can easily be drained by cystoduodenostomy. Large cysts which enlarge through and bulge inferiorly into the transverse mesocolon are best drained by cystojejunostomy with Roux-en-Y. Cysts located in the tail or body of the pancreas are technically more amenable to a resection (distal pancreatectomy and splenectomy) than cysts in the head and neck of the gland.

Although internal drainage into the upper gastrointestinal tract is the preferred method of treating pseudocysts, there are some exceptions. Infected or ruptured cysts or the acute cysts with thin, friable walls are best drained externally with wide-bore sump suction drains. In many instances, the resulting pancreatic fistula will gradually close spontaneously. Occasionally a second procedure is needed to implant the fistulous tract into the Roux-en-Y loop of jejunum. Two additional precautionary measures should be taken whenever a pancreatic cyst is drained:

1. The cyst fluid must be routinely sent for cytological examination and a representative sample of the cyst wall must be excised for histological examination. The injudicious drainage of a cystadenoma or a cystadenocarcinoma may thus be spotted and a planned reoperation for wide local excision entertained.
2. The cavity of the cyst must be explored with the index finger and septa gently divided. The cyst must be thoroughly irrigated prior to anastomosis with adjoining bowel or external drainage.

Percutaneous drainage of a pseudocyst under CT scan or ultrasound guidance is often advocated indiscriminately. It definitely has a temporizing role in some specific situations such as:

- a patient who is unfit for or who refuses an operation
- an acute pseudocyst which is rapidly enlarging and needs external drainage

- an infected pseudocyst which needs external drainage in a very sick septic patient
- a pseudocyst which is in an unusual location (e.g. mediastinum, pelvis) and is not readily amenable to internal drainage.

Alternative treatments for pancreatic pseudocysts

Although open surgical treatment remains the gold standard in the management of pancreatic pseudocysts, alterative techniques are used depending on the clinical state of the patient, the pathological anatomy of the pseudocysts and the available expertise. The alternative approaches include radiological (usually CT) percutaneous drainage which remains unpopular, endoscopic drainage which may be supplemented by EUS guidance and laparoscopic drainage.

Endoscopic drainage of pancreatic pseudocysts

Endoscopic drainage can be performed through the major papilla (transpapillary) or through the gastric (endoscopic cystogastrostomy) or duodenal walls (endoscopic cystoduodenostomy). The procedure may be performed with or without EUS guidance. Certain considerations determine the suitability of a particular patient for endoscopic cyst drainage. These include:

- site of cyst apposition to the gastric or duodenal wall
- determination of a site suitable for enterostomy
- presence of communication between the cyst and pancreatic duct or biliary duct
- cyst fluid consistency – solid material in the pseudocyst (detected by EUS) may preclude complete drainage and is an indication for surgery (open or laparoscopic)
- exclusion of malignancy.

Transpapillary drainage

Transpapillary drainage is successful only when the cyst is in communication with the pancreatic ductal system. Overall this is encountered in 60% and is more common with pseudocysts complicating chronic pancreatitis. Following documentation of the anatomy by the pancreatogram obtained during ERCP, a guide wire is passed through the pancreatic sphincter into the pseudocyst and pancreatic sphincterotomy performed. A stent is then passed over the guide wire preferably into the pseudocyst or to the site of ductal communication if the pseudocyst cannot be entered. A nasocystic drain can also be placed for additional drainage. A potential serious complication of transpapillary drainage is infection with abscess formation. Thus preprocedure antibiotics are used as prophylaxis together with preventive measures aimed at ensuring stent patency until the cyst drainage is achieved. EUS is advisable before transpapillary drainage to exclude solid debris from liquefied necrosis as this constitutes an absolute contraindication in view of the high infection rate in such cases. Transpapillary drainage has a success rate of 85%, with an average recurrence of 15% and morbidity of 8–12% (acute pancreatitis, infection).

Transenteric drainage without EUS guidance

This includes cystogastrostomy and cystoduodenostomy. Endoscopic cystoduodenostomy is nowadays preferred because of its increased safety, improved drainage and the relative dependency of the duodenum compared with the stomach for most pseudocysts. A luminal bulge by the cyst is required. In its absence, needle localization may require multiple passes with increased risk of complications. Thus it is important that the stomach or the duodenum share a common wall with the pseudocyst. CT scanning may be used for this, but EUS provides better confirmation of the contiguity of the wall between the cyst and the stomach or duodenum. Successful drainage is achieved in 90% with virtually no mortality. Complications of transenteric drainage include haemorrhage (6%), postprocedural pancreatitis, retroperitoneal perforation (3%) and infection or recurrence (6%) due to difficulty in maintaining an adequate fistula. There is virtually no mortality associated with the procedure.

EUS-guided endoscopic drainage

EUS provides important advantages in endoscopic drainage of pseudocysts, including accurate measurement of the distance between the enteric wall and the pseudocyst wall (distance >1 cm is considered a contraindication), detection of gastric varices associated with pancreatitis (contraindicates the procedure), imaging of gastric vessels with identification of pseudoaneurysms and detection of debris within pseudocysts which precludes effective drainage and increases the risk of infection. Nonetheless its superiority over the standard non-EUS-guided transenteric drainage remains to be confirmed by future studies.

Laparoscopic treatment of pancreatic pseudocysts

Laparoscopic treatment of pancreatic pseudocysts provides equivalent results to those of open surgery with the addition of the important clinical advantages in terms of accelerated postoperative recovery and reduced hospital stay. There are two techniques: laparoscopic infracolic cystenterostomy (Figure 27.18a–c) or transgastric stapled cystogastrostomy (Figure 27.18d, e). The transgastric cystogastrostomy is greatly facilitated by the hand-assisted laparoscopic surgery approach.

In a systemic review based on 19 laparoscopic and 25 endoscopic publications on the results of patients who underwent 118 and 583 laparoscopic and endoscopic drainage procedures, respectively, the pancreatic pseudocysts were considerably larger in the laparoscopic series (mean 13 vs 7 cm). Despite this, successful resolution of the pseudocyst was higher in the laparoscopic series (98%) than in the endoscopic series (81%), with morbidity rates of 4% and 12% and mortality rates of 0% and 0.4%, respectively. In the follow-up period, which was however significantly shorter in the laparoscopic series (mean 13 vs 24 months), the recurrence rates were 2.5% in the laparoscopic series and 14.4% for the patients treated endoscopically. The reintervention rate was also lower in the laparoscopic series (1% vs 12%). Obviously longer follow-up studies are needed to confirm the superiority of the laparoscopic approach.

Management of pancreatic abscess/infected pancreatic/peripancreatic necrosis

The term 'pancreatic abscess' should not be used to connote a localized collection of pus in the lesser sac (infected pancreatic

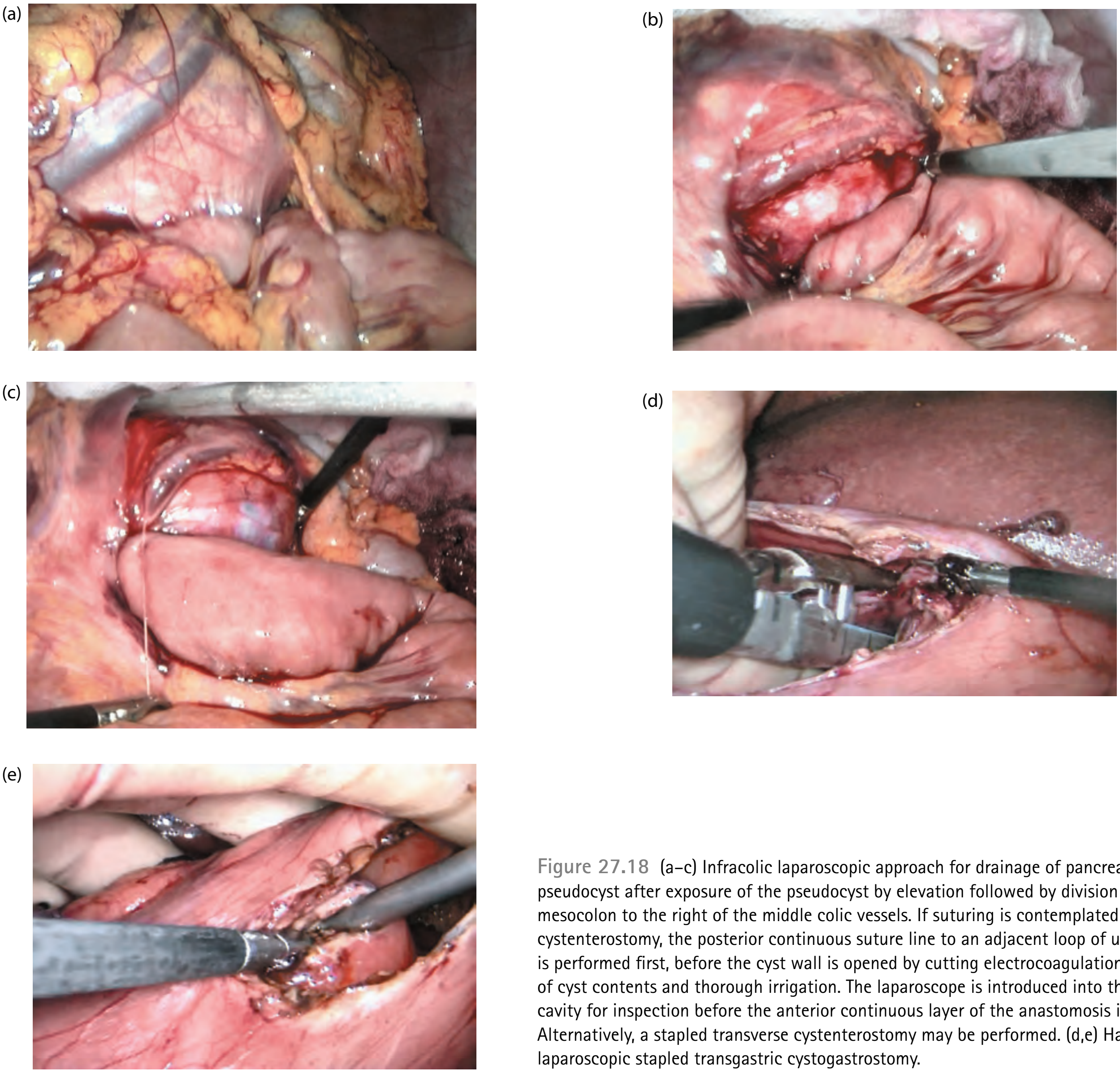


Figure 27.18 (a–c) Infracolic laparoscopic approach for drainage of pancreatic pseudocyst after exposure of the pseudocyst by elevation followed by division of the mesocolon to the right of the middle colic vessels. If suturing is contemplated for the cystenterostomy, the posterior continuous suture line to an adjacent loop of upper jejunum is performed first, before the cyst wall is opened by cutting electrocoagulation, aspiration of cyst contents and thorough irrigation. The laparoscope is introduced into the cyst cavity for inspection before the anterior continuous layer of the anastomosis is completed. Alternatively, a stapled transverse cystenterostomy may be performed. (d,e) Hand-assisted laparoscopic stapled transgastric cystogastrostomy.

pseudocyst) which is easily drained externally with good results. A pancreatic abscess implies the presence of extensive pancreatic and peripancreatic necrosis with secondary infection. Nowadays the term IPN (infected pancreatic necrosis) is used to avoid confusion. IPN carries a very high mortality and morbidity and usually occurs after the second week of the onset of pancreatitis. IPN is always accompanied by organ failure and other systemic complications and may be further complicated by massive haemorrhage.

Infection of the pancreatic necrosis is thought to arise by translocation of enteric bacteria probably from the transverse colon. The most commonly isolated bacteria in IPN are *Escherichia coli, Klebsiella pneumoniae, Enterococcus faecalis, Staphylococcus aureus, Pseudomonas aeruginosa, Proteus mirabilis* and *Streptococcus* species, and the infection is often mixed. It is important to stress that non-IPN diagnosed by CT is not an indication for surgery and necrosectomy unless the condition of the patient is deteriorating. Evidence-based guidelines for the management of IPN include:

- CT-guided or ultrasound-guided FNA for bacteriology should be performed in patients suspected of having IPN
- IPN accompanied by signs of sepsis is an indication for surgical intervention
- patients with sterile pancreatic necrosis should be managed conservatively, and surgical intervention only performed in selected cases, e.g. those with persistent organ failure or severe clinical deterioration despite maximum intensive care
- early surgical intervention (before third or fourth week) for IPN is not usually indicated as delay enables softening and demarcation
- necrosectomy with drainage is recommended as the surgical treatment for IPN
- drainage after necrosectomy is essential
- percutaneous drainage may be performed for pancreatic abscess but if this fails to improve the clinical condition, immediate surgical drainage is necessary.

The role of appropriate prophylactic antibiotic therapy remains controversial but has gained support in clinical practice.

The choice of antibiotics is based on the expected microbial flora and the extent of antibiotic penetration into the abscess/necrotic tissue. Although case series have been reported of patients treated successfully without surgical intervention but with CT-guided drainage, this is not the orthodox recommended treatment, although it is indicated in patients who cannot tolerate an open procedure. Thus surgery is the mainstay of treatment for IPN but several unresolved issues remain, including selection of patients for surgery, optimal timing for the intervention and the technique. The absence of infection is not an absolute contraindication, as although the vast majority of these patients can be successfully treated without surgical intervention there is a small subset of patients (10%) in whom extensive necrosis warrants surgery because of deteriorating organ failure despite maximal support or persisting symptoms which preclude hospital discharge despite several weeks of conservative treatment.

The timing of surgery is an important determinant of outcome. Early surgery (within the first week) is avoided in view of the reported high mortality. Surgery for IPN should be undertaken in the 3–4 weeks after onset when demarcation and softening of the necrosis has taken place, thus minimizing the risk of bleeding during the necrosectomy. The aim of operation is to remove the necrotic tissue and to provide adequate drainage for the remaining debris while preserving viable pancreatic tissue.

The standard recommended approach is by open laparotomy, although other alternative less invasive approaches are currently being evaluated in specialist centres:

- laparoscopic infracolic transmesocolic necrosectomy with closed peritoneal drainage and irrigation (Buerger's irrigation)
- laparoscopic transgastric necrosectomy with external drainage
- retroperitoneal necrosectomy and drainage with endoscopic (video-assisted) or radiological guidance.

To date, the reported experience with these minimal access techniques has been limited to a few centres and, although the reported results have been very promising in terms of reduced mortality in these critically ill patients, selection of cases is important and, for the time being, the general consensus is that these minimal access approaches should be confined to specialist centres. The experience in Dundee, UK, with the laparoscopic infracolic necrosectomy and closed peritoneal drainage and irrigation (Figure 27.19) has been positive in reducing mortality though it has not reduced the hospital stay of the surviving patients.

Thus delayed open surgery remains the gold standard treatment for IPN, although the technique is not standardized and different centres use one of the following:

- necrosectomy combined with open packing
- planned staged relaparotomies with repeated lavage
- necrosectomy with closed continuous lavage of the retroperitoneum (Beger's technique)
- necrosectomy with closed packing.

The reported series however indicate that necrosectomy with closed continuous lavage of the retroperitoneum or closed packing seem to be associated with a lower morbidity.

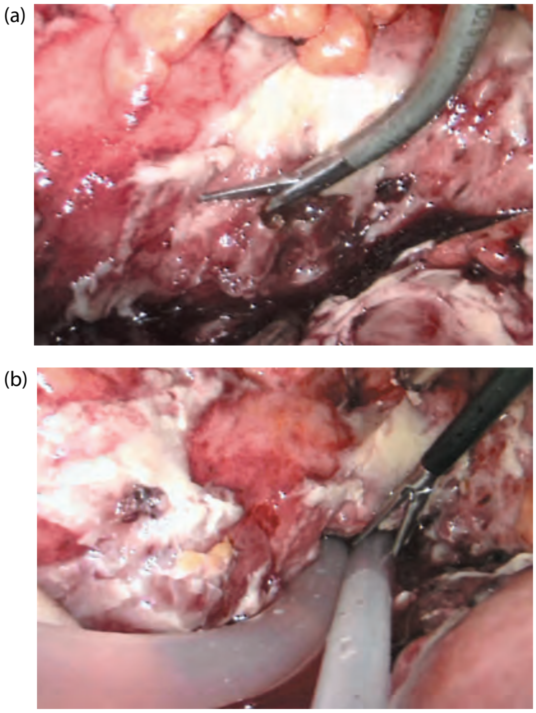


Figure 27.19 (a,b) The infected pancreatic necrosis is approached infracolically after elevation of the transverse colon and division of the transverse mesocolon to the right of the middle colic vessels. Only a small hole (enough to admit a 5 mm suction device) is made and the infected liquid contents aspirated before pulsed irrigation and extension of the opening for the necrosectomy which is stopped as soon as oozing of blood is detected. Two large drains are inserted for postoperative irrigation and drainage with hypertonic dialysate solution.

Recurrent pancreatitis

Any patient who has recovered from one or more attacks of pancreatitis must be investigated to identify and, if possible, to eliminate the aetiological factors. The need to identify surgically remediable problems, most commonly gallstones, is obvious. Ultrasonography of the biliary tract can be used even in the acute stage.

Stenosis of the sphincter of Oddi, often called papillitis, is without doubt a rare cause of recurrent pancreatitis. The stenosis is rarely of a primary nature and is probably due to a temporary impaction and later passage of a gallstone. Confidence in the diagnosis is strengthened by:

- documented episodes of biliary obstruction or cholangitis
- the endoscopic visualization of an inflamed papilla
- demonstrated tightness of the ampullary orifice when it is cannulated endoscopically
- documentation of delay in emptying of the common bile duct
- intermittent elevations of liver enzymes, especially serum alkaline phosphatase
- positive morphine–prostigmine (Nardi) test – reproduction of pain and enzyme elevation
- documentation of elevated pressures in the common bile duct by ERCP.

Treatment is sphincteroplasty, which includes division of the sphincter of Oddi and the septum between the common bile duct and the pancreatic duct to relieve both biliary and pancreatic outflow obstruction. The situation should be suspected in patients who have had more than one attack of pancreatitis and/or cholangitis without any other discernible cause. One word of caution – the condition of 'primary papillitis' has been overdiagnosed without adequate documentation of unexplained abdominal pain which is often not even of pancreatic or biliary origin. In such situations, the indiscriminate use of sphincteroplasty, often wrongly performed, has led the operation into disrepute. Endoscopic papillotomy has a definite place when the biliary sphincter alone needs to be divided.

Congenital malformations in the pancreas or duodenum around the ampullary region can cause recurrent pancreatitis in both children and adults. EUS and ERCP are invaluable in delineating these abnormalities, some of which may be amenable to surgical correction. Probably the commonest congenital abnormality leading to recurrent pancreatitis is the pancreas divisum, which is an anatomical variant occurring when there is failure of the two embryonic pancreatic ductal systems to unite. In this situation, the duct of Wirsung is very small and may measure no more than 1–2 cm in length while the duct of Santorini becomes the major ductal drainage system of the pancreas and maintains its communication with the duodenum through the minor papilla. Following secretin administration, large volumes of juice can be visualized endoscopically from the minor papilla with little or none coming from the main papilla or aspirated from the duct of Wirsung during cannulation. The high incidence of recurrent pancreatic pain in patients with pancreas divisum may well be due to the very small papilla of the duct of Santorini which, in these patients, drains the majority of the pancreas, creating a marked relative stenosis of the ampulla.

The relief of pain by all treatment modalities is unsatisfactory leading one to question the proposed pathophysiological explanation. Endoscopic papillotomy, with or without stenting of the duct of Santorini, and surgical transduodenal sphincteroplasty of both the minor and major papilla have not yielded favourable long-term results. Eventually, the recurrent pancreatitis coupled with the trauma induced by repeated endoscopic and/or surgical approaches force the surgeon to consider a Whipple pancreatoduodenectomy or a duodenum-preserving proximal pancreatectomy for pain relief.

Left-sided (sectorial) portal hypertension due to pancreatitis

This occurs from occlusion by compression or by thrombosis of the splenic vein and is discussed elsewhere (Chapters 24 and 26).

Acute pancreatitis in children

In recent years, acute pancreatitis is being recognized with greater frequency in infancy and childhood. Although the treatment is not substantially different from that of adults, the aetiological factors are totally different. Biliary tract disease is not usually an aetiological factor; nor is alcoholism. A large number of mild acute pancreatitis results from viral infections such as mumps. Blunt abdominal trauma, resulting from relatively minimal injury, is probably the most common cause. Unlike the situation in adults, isolated pancreatic injury is commonly reported. A trifling fall upon a toy, such as a tricycle handlebar, may result in pancreatic trauma sufficient to induce a traumatic pancreatitis. Two clinical pictures emerge: (1) an acute abdominal emergency necessitating exploratory laparotomy when the pancreas is found to be inflamed – in such situations, care must be taken to exclude a complete pancreatic transection or major ductal injury; (2) the initial symptoms are mild and often the child is not even taken to the hospital for several days or weeks when a pancreatic pseudocyst or pancreatic ascites has developed.

Drug-induced pancreatitis, familial pancreatitis (usually associated with hyperlipidaemia or amino aciduria) and calculous disease of the biliary tree are all uncommon in children, although occasional cases with pigment stones due to congenital spherocytosis have been described. Obstruction of the ampulla of Vater due to a congenital anomaly of the pancreatic ductal ampulla has also been documented and has been corrected by an adequate sphincteroplasty. Roundworms entering into the pancreatic duct and causing pancreatitis have also been reported from time to time. Duodenal obstruction due to an annular pancreas may lead to current acute pancreatitis which responds to duodenal decompression.

The management of acute or recurrent acute pancreatitis in children is essentially the same as in adults. The prognosis is much better even in the presence of complications. If the aetiological factors are removed, recovery is complete. Acute or recurrent acute pancreatitis does not progress into the chronic or chronic relapsing form in the vast majority of children.

Chronic pancreatitis

Chronic pancreatitis is defined as a continuing inflammatory disorder of the pancreas characterized by irreversible pathological changes which cause abdominal pain and/or permanent impairment of pancreatic exocrine and endocrine function. The disease may present clinically with an individual symptom or a combination of symptoms but by far the commonest presentation is with pain, which may be either intermittent or chronic and persistent, leading to regular narcotic opiate intake and addiction. The natural history of chronic pancreatitis is characterized by progressive damage of the pancreatic parenchyma with various degrees of exocrine and endocrine pancreatic insufficiency becoming clinically manifest with time: malabsorption and diabetes mellitus. The progressive fibrosis may also lead to large bile duct obstruction and deepening jaundice. Chronic pancreatitis increases significantly the risk for the development of pancreatic cancer, and in jaundiced patients with established disease, differentiation between chronic pancreatitis and ductal adenocarcinoma of the pancreas may be extremely difficult despite intensive investigations including pancreatic biopsy. Not infrequently, the exact pathology in the individual patient can only be established after pathological examination of the resected specimen following a pancreaticoduodenectomy.

Classification

Despite considerable research, there is as yet no internationally agreed classification of chronic pancreatitis that unifies the clinical features, the varied aetiology and radiological appearance. Some such as the Marseille classification and its various updates over the years distinguish the morphological, functional and clinical features of three types: (1) acute relapsing pancreatitis, (2) chronic pancreatitis and (3) chronic relapsing pancreatitis. The Cambridge classification recognizes two types as distinct forms of the disease: (1) chronic calcifying pancreatitis and (2) chronic inflammatory pancreatititis.

Perhaps the most useful is the TIGAR-O, which classifies chronic pancreatitis on aetiology (**T**oxic and metabolic, **I**diopathic, **G**enetic, **A**utoimmune and **R**ecurrent and severe acute pancreatitis, associated chronic pancreatitis, and **O**bstructive). The TIGAR-O classification has certain merits: (1) it includes a group of patients with genetic disorders which predispose to chronic pancreatitis induced by various agents, and (2) it considers the subset of patients with severe acute pancreatitis with severe parenchymal damage leading to chronic disease. Even so, this classification fails the clinical surgeon in staging the severity of the disease and indicating the morphological type (dilated duct as distinct from small duct disease) which in the end dictates the type of surgical intervention, if and when this becomes indicated.

The various categories of the TIGAR-O classification (version 1.0) are listed below.

1. Toxic–metabolic: alcoholic, tobacco smoking, hypercalcaemia, hyperparathyroidism, hyperlipidaemia, chronic renal failure, medications (phenacetin abuse), toxins (organotin compounds, e.g. DBTC).
2. Idiopathic: early onset, late onset, tropical, tropical calcific pancreatitis, fibrocalculous pancreatic diabetes.
3. Genetic: autosomal dominant – cationic trypsinogen (codon 29 and 122 mutations); autosomal recessive/modifier genes – *CFTR* mutations, *SPINK1* mutations; cationic trypsinogen (codon 16, 22, 23 mutations), α_1-antitrypsin deficiency.
4. Autoimmune: isolated autoimmune chronic pancreatitis (AIP), syndromic AIP, Sjögren syndrome-associated chronic pancreatitis, inflammatory bowel disease-associated chronic pancreatitis, primary biliary cirrhosis-associated chronic pancreatitis.
5. Recurrent and severe acute pancreatitis: postnecrotic (severe acute pancreatitis) recurrent acute pancreatitis, vascular diseases/ischaemic, postirradiation.
6. Obstructive: pancreatic divisum, sphincter of Oddi disorders, duct obstruction (e.g. tumour), preampullary duodenal wall cysts, post-traumatic pancreatic duct scars

Especially in Western countries, alcohol abuse is by far the most common cause of chronic pancreatitis, although only 10–20% of chronic alcoholics develop acute or chronic pancreatitis. This observation indicates the interaction of genetic or other environmental cofactors in the pathogenesis of alcoholic pancreatitis, and, indeed, an increased prevalence of genetic mutations known to produce chronic pancreatitis has been documented in some of these patients. Smoking is associated with increased risk for chronic pancreatitis which is independent of alcohol use. The constituents of tobacco smoke decrease pancreatic secretion, induce oxidative stress and increase the development of pancreatic calcification. Drug-induced pancreatitis is much more usually acute rather than chronic pancreatitis.

The important metabolic conditions associated with acute and chronic pancreatitis are (1) hypercalcaemia and (2) chronic renal failure. Severe hypercalcaemia can cause acute pancreatitis through trypsin-mediated mechanisms. Persistent untreated hypercalcaemia will ultimately cause chronic pancreatitis as evidenced by the established association between hyperparathyroidism and chronic pancreatitis. The most likely mechanism is thought to be recurrent acute pancreatitis progressing to chronic disease (necrosis–fibrosis theory). Another mechanism that has been suggested is protein plug formation by the hypercalcaemia (obstructive theory). The prevalence of acute and chronic pancreatitis is increased in patients with renal failure although the pathological reason for this association remains unknown, but has been variously attributed to (1) uraemic toxicity, (2) recurrent volume contraction during haemodialysis, (3) recurrent acute pancreatitis from secondary hyperparathyroidism and (4) alteration of the gastrointestinal hormone profile causing pancreatic exocrine dysfunction.

It is likely that the idiopathic category of chronic pancreatitis will ultimately be dropped as a result of progress in genomic studies identifying defect genes which predispose to the postnatal development of chronic pancreatitis following exposure to environmental and immune-mediated risk factors. Currently, idiopathic pancreatitis accounts for 10–30% of patients and is classified as early and late onset, with early-onset idiopathic chronic pancreatitis presenting in the first two decades of life with severe abdominal pain, and pancreatic insufficiency developing much later after several years. In contrast, late-onset idiopathic chronic pancreatitis, encountered in the fourth or fifth decade, presents with minimal pain, but with established pancreatic insufficiency at the time of diagnosis. Exocrine and endocrine insufficiency and pancreatic calcifications are much more commonly encountered in late-onset idiopathic chronic pancreatitis. The serine protease inhibitor Kazal type 1 (*SPINK1*) mutation has been found in many patients with early idiopathic chronic pancreatitis. The *SPINK1* gene encodes for the pancreatic secretory trypsin inhibitor, which normally counteracts the effects of activated trypsin. Mutations causing loss of function of this protein thus increase the risk of development of acute and chronic pancreatitis.

AIP also known as non-alcoholic duct-destructive chronic pancreatitis is rare, and many cases are labelled mistakenly as idiopathic chronic pancreatitis. AIP may occur in isolation or in association with other autoimmune disorders, e.g. Sjögren syndrome, primary sclerosing cholangitis, inflammatory bowel disease, etc. The clinical features of AIP include minimal pain, hypergammaglobulinaemia, autoantibodies (antinuclear antibody, antilactoferrin, anticarbonic anhydrase I and II, antismooth muscle), diffuse enlargement of the pancreas on CT imaging, absence of calcifications and cyst pseudocyst formation and marked improvement with steroid therapy. The histology of AIP shows diffuse fibrosis with infiltration of lymphocytes and plasmacytes, especially surrounding the pancreatic duct. ERCP/MRCP shows significant narrowing and irregularity of the pancreatic ducts. The pathogenesis of AIP is thought to result from immune-mediated damage to the pancreatic ductal cells. The cell infiltrates around the duct exhibit a distinct profile consisting of increased numbers of $CD4^+$ and $CD8^+$ T-cells.

Even one very severe episode of acute pancreatitis may result in permanent pancreatic damage with glandular fibrosis and hypofunction leading to chronic pancreatitis, but more commonly recurrent acute pancreatitis from any cause is responsible for the development of chronic pancreatitis through the necrosis–fibrosis pathway. The exceptions to this seem to be recurrent gallstone or hypertriglyceridaemia-associated pancreatitis, where progress to chronic pancreatitis is rare.

Obstructive pancreatitis is a separate category in the TIGAR-O and Marseille systems. Experimentally, obstruction of the main pancreatic duct produces changes of chronic pancreatitis within weeks in several animal models. The pathological features of obstructive pancreatitis in humans include uniform inter- and intralobular fibrosis and marked destruction of the exocrine parenchyma in the territory of obstruction, with absence of plug formation and calcifications. Pancreatic tumours (pancreatic adenocarcinoma, neuroendocrine tumours and intrapapillary mucinous tumours) can produce both recurrent acute and chronic pancreatitis as a result of duct obstruction. Obstruction of the main pancreatic duct leads to inspissation of the pancreatic juice which becomes lithogenic (with stone formation) and induces recurrent episodes of acute inflammation with periductular fibrosis. The pancreatic intraductal pressure is raised (ductal hypertension) and this is responsible for the pain of obstructive chronic pancreatitis and itself promotes fibrosis and glandular damage. In patients with large ducts, chronic hypertension results from stone and stricture formation. These patients require surgical or endoscopic decompression, which relieves the pain. Experimental studies have indicated that the pancreatic ductal hypertension is accompanied by reduced pancreatic blood flow and this is thought to play a role in the progression of fibrosis of the gland.

Surgical treatment of chronic pancreatitis

Maintenance of adequate nutrition, enzyme replacement and/or insulin supplements may be necessary in the management of exocrine and/or endocrine insufficiencies. The input of social services and of an interested psychiatric team is essential to manage drug addiction and alcoholic problems which are often present. Direct operative procedures on the parenchyma of the gland and/or its ductal system are indicated almost exclusively for the relief of pain. The limits and hazards of surgical treatment of these patients must be emphasized. No surgical procedure can restore either the endocrine or exocrine function of the pancreas. Nor is it likely to prevent further loss of glandular function. The conversion of a non-reformed alcoholic or drug addict into an insulin-dependent diabetic by major pancreatic resection is likely to be lethal and must be avoided. Rehabilitation of the patient must be planned well in advance otherwise surgical intervention for pain is doomed to failure. The life expectancy of the non-reformed alcoholic drug addict is extremely limited and is often shortened by the complications and late sequelae of operations. Avoidance of alcohol is a more important determination of outcome after operation than the type of procedure performed.

There are two indications for surgical treatment:

1. *Intractable pain* – it is crucial to delineate the frequency, persistence and degree of pain. The decision to advise operation is influenced by several factors including the degree of disruption of the patient's life, the narcotic need, the control of alcoholism, the age and general condition of the patient and, often, the surgeon's personal preferences.
2. *The development of complications* – these include (1) lower bile duct obstruction; (2) duodenal obstruction; (3) vascular involvement; (4) pancreatic cysts, pseudocysts, pancreatic ascites, and pleural effusions; and (5) the presence of a dominant mass leading to the fear or suspicion of cancer.

It is important to emphasize that correction of a complication does not invariably relieve any associated pain.

Lower bile duct obstruction

The lower portion of the common bile duct passes through the head of the pancreas and is at risk of being narrowed by inflammation and fibrosis in this region. If frank obstructive jaundice is present, the onus is on the surgeon to exclude preoperatively and operatively the presence of an underlying cancer. On occasions, this may only be possible after a total pancreatoduodenectomy. More commonly, the patient has low-grade cholangitis and pain indistinguishable from pancreatic pain. Frank suppurative cholangitis and secondary biliary cirrhosis have also been described. In the mild case, serum alkaline phosphatase elevation is the most consistent although non-specific effect of biliary obstruction. As a rule, investigation of the biliary tree (by MRCP/ERCP) is mandatory whenever surgical treatment of chronic pancreatitis is being considered. Relief of the partial common bile duct obstruction is, in some instances, all that is needed to relieve the pain.

Duodenal obstruction

This rarely occurs in patients with severe chronic pancreatitis and enlargement of the head of the pancreas. Here again a concomitant pancreatic cancer must be excluded by appropriate biopsies (in the young patient) or by pancreatoduodenctomy (in the older patient). Vagotomy and gastrojejunostomy will adequately relieve the duodenal obstruction.

Development of vascular complications

These include multiple pseudoaneurysms and sectorial portal hypertension. These have been discussed in the section on acute pancreatitis.

Once a decision for surgical treatment has been made, the two most important preoperative investigations are ERCP and angiography. ERCP will delineate the state of the main pancreatic duct and common bile duct and helps in the planning of surgical therapy. However, the findings on ERCP often do not correlate with the patient's symptoms. ERCP does not indicate the state of the parenchyma, nor does it indicate the need for operative intervention except if cytological examination of the pancreatic duct aspirate is positive for cancer. However, ERCP provides a good indication of the choice of operation. Similarly, angiography delineates the anatomy of the foregut vasculature as well as vascular complications which may necessitate an alteration in surgical strategy.

Although ERCP is often used as the gold standard for defining chronic pancreatitis, it is an invasive test that carries the risk of cholangitis and acute pancreatitis (approximately 4%). Angiography is also invasive and usually reserved for therapeutic embolization in cases of bleeding. Therefore, it is not surprising that MRCP and EUS are redefining the indications for ERCP and angiography in the evaluation of chronic pancreatitis. EUS allows viewing of the pancreas through the duodenal and gastric wall with high-frequency ultrasound probes, which are capable of much greater resolution of fine structural detail than conventional abdominal ultrasound. In addition, intestinal gas does not provide a barrier to the pancreas. Multiple criteria for the diagnosis of chronic pancreatitis have been proposed, including parenchymal changes described as hyperechogenic foci, hyperechogenic stranding, lobularity of the gland and cyst formation. Ductal changes include hyperechoic thickening, irregularity, dilatation, visible side branches and calcified duct stones. The EUS technology has the limitations of being operator dependent, and therefore the exact role of its use in the diagnosis of chronic pancreatitis is not yet established as this requires long-term longitudinal studies and careful follow-up of patients with suspected mild disease.

As a general rule, surgery should be conservative when there is no endocrine or exocrine insufficiency and a dilated duct of Wirsung is present. Dilatation (diameter >3 mm) of the main pancreatic duct, with or without partial stenosis of the duct, at a number of points producing a 'chain of lakes' appearance may be associated with pancreatic stones in the duct. In this situation, longitudinal filleting of the main pancreatic duct and side-to-side anastomosis to a Roux-en-Y loop of the jejunum (modified Puestow operation) is highly appropriate after removing any stones if present. Relief of pain is accomplished in about 70% of patients who stop consuming alcohol, although recurrence of pain is common after variable intervals. Surgery should be radical when pancreatic cancer is suspected and/or cannot be excluded, there is established endocrine and exocrine insufficiency and there is extensive pancreatic destruction by ductal sclerosis, glandular fibrosis, calcification and multiple pseudocysts.

A single pseudocyst may be drained internally as a preliminary step to see if the patient's pain is relieved. The presence of multiple cysts or the reformation of cysts is an indication for pancreatic resection. The 95% distal pancreatectomy (Child's procedure) is not uniformly successful, and recurrent pain associated with pancreatitis in the region of the head and the uncinate process then necessitates a second-stage pancreatoduodenectomy which can be technically difficult and hazardous. When pancreatic cancer is not suspected, total pancreatectomy can be performed in one stage with preservation of the whole stomach, pylorus and first part of the duodenum by careful preservation of the blood supply to the pyloroduodenal area. This diminishes postoperative problems associated with reduced gastric reservoir capacity and dumping syndrome.

Because 40–60% of patients with painful chronic pancreatitis exhibit a ductal ectasia, decompression of the pancreatic ductal system has become one of the main therapeutic principles, based on the established association between ductal ectasia and intraductal hypertension. Many different approaches to decompressing the pancreatic duct have been described. In 1956, Puestow and Gillesby described a technique in which drainage of the main pancreatic duct was accomplished by performing a longitudinal laterolateral pancreaticojejunostomy after resection of the pancreatic tail and splenectomy. Although the procedure met with some success, 15–40% of patients did not experience permanent pain relief. In an effort to improve results with drainage alone, several surgeons, including Beger and Frey, have combined resection with drainage. The Beger procedure includes a subtotal resection of the pancreatic head following transection of the pancreas anterior to the portal vein. The gastroduodenal passage and common bile duct continuity are preserved. The body of the pancreas is drained by an end-to-end or end-to-side pancreaticojejunostomy using a Roux-en-Y loop. The Frey procedure differs from the Beger in that there is no transection of the pancreas in front of the portal vein. For reconstruction, a longitudinal pancreaticojejunostomy is used draining the resection cavity of the head, body and tail of the pancreas.

A less radical approach which has been reported to be occasionally successful is the performance of a truncal vagotomy, antrectomy and gastrojejunal (Billroth II or Pólya) anastomosis. The rationale of this operation is the elimination of neural and hormonal stimuli to pancreatic secretion, especially those normally triggered by eating. No convincing data are available to support these contentions. Two other operations, namely cholecystectomy (for established gallbladder disease) and parathyroidectomy (for proved hyperparathyroidism), are sometimes advocated to reduce the severity of chronic pancreatitis. The incidence of gallstones in patients with chronic pancreatitis is the same as that in the general population. Cholecystectomy should be advised based on symptoms of gallbladder disease and on the risk of complications. It will not affect the natural history of chronic pancreatitis. Similarly, hyperparathyroidism should be treated to avoid the sequelae of severe hypercalcaemia without influencing the course of any incidental chronic pancreatitis.

Splanchnic neurectomies and coeliac ganglion block have generally been disappointing in the control of chronic pancreatic pain.

First reported in 1943, splanchnicectomy for the management of intractable pancreatic pain was practically forgotten because of the invasiveness required (laparotomy or thoracotomy in patients with limited survival) and the inconsistent results achieved. With the evolution of minimal access surgery, however, interest has been rekindled. The first thoracoscopic splanchnicectomy for pancreatic cancer pain was performed in 1993 and was soon followed by numerous other reports advocating its use for chronic pancreatitis pain. In this procedure, four trocars are optimal: camera, lung retraction and two working ports.

After transecting the inferior pulmonary ligament, the lung is retracted anteromedially. The sympathetic trunk is identified as a guide to the greater splanchnic nerve, which lies medial to it close to the aorta on the left and the oesophagus on the right. The overlying pleura is incised and the nerve is dissected free and transected sharply. Left-sided splanchnicectomy alone seems effective in most patients. Preliminary studies indicate that the procedure is effective, but long-term follow-up is lacking at the present time. Promising results have been obtained recently with thoracoscopic bilateral splanchnicectomy although the reported experience is limited and the follow-up short.

Neoplasms of the non-endocrine pancreas

Benign neoplasms of the non-endocrine pancreas are exceedingly rare and are of no clinical significance unless they become large enough to be palpable or to impinge on adjacent structures (common bile duct, duodenum, stomach or main pancreatic duct) and cause symptoms. Since both solid and cystic benign tumours are rarely found at laparotomy or at necropsy, there is no evidence to suggest that they represent an early phase in the development of the more common malignant neoplasms. The reported benign tumours of the non-endocrine pancreas include adenoma, cystadenoma, lipoma, fibroma, leiomyofibroma, myoma, haemangioma, lymphangioma, haemangioendothelioma and neuroma. These diagnoses should only be made after exclusion of the more frequent malignant tumours by some form of representative (preferably excisional) biopsy.

Pancreatic cancer

The term ‘pancreatic cancer’ is sometimes used to include all types of malignant neoplasms of the non-endocrine pancreas (Box 27.3) as well as malignant islet cell tumours (Table 27.6).

In clinical practice, pancreatic cancer is synonymous with PDAC (arising in the exocrine pancreas), which constitutes 90% of all primary malignant tumours arising from the gland. When the cancer originates in the head of the pancreas (in about 70% of cases) it must also be differentiated from cancer arising from the ampulla, duodenum or lower common bile duct, which has a much better prognosis than true pancreatic adenocarcinoma. The incidence of pancreatic cancer has tripled over the past 40 years throughout the West. It is highly fatal and has one of the lowest 5 year survival rates (1–2%) of all cancers. About 29 000 new pancreatic cancers are diagnosed each year in the USA and the disease now accounts for 10% of all the cancers of the digestive tract (second behind colorectal cancer). It is the fourth most common cancer of all sites as a cause of death (behind lung, colorectal and breast).

Cancer of the pancreas is distinctly more common in older people and is relatively uncommon, but not altogether rare, below the age of 55 years. It occurs more frequently in men than in women but the male-to-female ratio has decreased in recent years, suggesting that more women are now being diagnosed with this cancer. The exact causative factors responsible for the increase in incidence of pancreatic cancer are unknown. A high-protein and high-fat diet, characteristic of the Western population, has been implicated epidemiologically as a possible factor. The strongest association is between pancreatic cancer and cigarette smoking. Exposure to industrial carcinogens, especially betanaphthylamine and benzidine, has been documented in pancreatic cancer patients. A higher than normal incidence rate of the neoplasm has also been reported in chemists, workers in metal industries, and coke and gas plant employees. In interpreting these retrospective epidemiological data, it must be remembered that the general class of ‘labourer’ has, both in the USA and in Britain, an excessively high incidence of pancreatic cancer so that occupational risk is mixed with social class risk. Industrial causes of pancreatic cancer are probably less important than is believed, or the causative exposures are far more widespread in most occupations than is generally accepted.

BOX 27.3 Pathological classification of primary malignant neoplasms of the pancreas (non-endocrine)

1. Duct (ductular) cell origin 90%
- Duct cell adenocarcinoma
- Giant cell carcinoma
- Giant cell carcinoma (epulis-osteoid)
- Adenosquamous carcinoma 10%
- Microadenocarcinoma
- Mucinous (colloid) carcinoma
- Intraduct papillary mucinous neoplasms

2. Acinar cell origin <1%
- Acinar cell carcinoma
- Cystadenocarcinoma (acinar cell)

3. Connective tissue origin <1%
- ‘Osteogenic’ sarcoma
- Leiomyosarcoma
- Haemangiopericytoma
- Malignant fibrous histiocytoma

4. Uncertain histogenesis 8%
- Pancreaticoblastoma
- Papillary and cystic neoplasm
- Mixed type: duct and islet cells
- Unclassified

5. Miscellaneous others <1%
- Malignant melanoma
- Oncocytoma
- Neuroblastoma
- Plasmacytoma
- Lymphoma

Molecular biology of pancreatic cancer

Sporadic cancers of the pancreas are frequently associated with the activation of an oncogene, K-*ras*, and the inactivation of multiple tumour suppressor genes, including *p53*, *DPC4*, *p16* and *Rb*.

Oncogenes of the *ras* family (e.g. H-*ras*, N-*ras* and K-*ras*) are among the most common activated oncogenes found in

Table 27.6 Islet cell tumours and associated clinical syndromes

Cell type	Peptide product	Tumour	Clinical picture
A cell	Glucagon	Glucagonoma	Diabetes, necrolytic migratory erythema, stomatitis, glossitis Weight loss, weakness
B cell	Insulin	Insulinoma	Neuroglycopenia Clouded sensorium Behaviour disturbance Seizures Transient neurological deficit Adrenaline discharge Sweating Tremulousness Palpitation Hunger
D cell	Gastrin	Gastrinoma	Fulminant peptic ulceration Diarrhoea Malabsorption
D-cell	Somatostatin	Somatostatinoma	Diabetes, pancreatic malabsorption with steatorrhoea, gallstones Weight loss
D_1-cell?	Vasoactive intestinal polypeptide	VIPoma	Fulminant diarrhoea, hypokalaemia, hypercalcaemia, diabetes, flushing
D_1-cell?	Human pancreatic polypeptide	HPPoma	Uncertain
9	ACTH	?	Cushing syndrome
?	Hydroxyindole	Carcinoid	Carcinoid syndrome?
?	Prostaglandin	?	Same as VIPoma?
9	ADH	?	?
?	None	Non-functioning islet cell tumour	Same as exocrine tumour but slow growing

ACTH, adrenocorticotrophic hormone; ADH, antidiuretic hormone.

human cancer. The K-*ras* gene, on chromosome 12, encodes a 2.0 kb transcript that is highly conserved across species and is translated into the p21–ras protein. These proteins are located in the plasma membrane and transduce growth and differentiation signals from activated receptors to protein kinases within the cell. Up to 90% of human pancreatic cancers have *ras* gene mutations, most of which are K-*ras* mutations and these seem to be an early event in pancreatic carcinogenesis.

The *p53* gene is the most frequently mutated gene identified in human cancers. The *p53* tumour suppressor gene encodes a 393 amino acid phosphoprotein that acts as a potent transcription factor regulating many cellular functions and has been called the ‘guardian of the genome’ because its role is to induce apoptosis (cell death) when an abnormality cannot be repaired. It inhibits oncogene-induced transformation, blocks progression through the G1 phase of the cell cycle, modulates the expression of growth control genes and maintains genome stability. Mutations in the *p53* gene result in an abnormal protein unable to carry out its regulatory functions, leading to cell transformation and neoplasia. A number of studies have indicated that *p53* mutations are relatively common in adenocarcinoma of the pancreas, occurring in 70% of patients. In addition, *p53* mutation may be an independent prognostic factor in patients with PDAC.

The *pl6* gene, which is located on chromosome 9p, encodes a protein that binds cyclin D–cdk4 complexes. When *pl6* binds to these complexes, it inhibits the phosphorylation of a number of growth and regulatory proteins, including the retinoblastoma protein (Rb). Hypophosphorylated Rb binds to and sequesters transcription factors that would otherwise promote the Gl/S transition in the cell cycle. Inactivation of *pl6* therefore inactivates another important cell cycle checkpoint. Mutations in the *p16* gene are found in approximately 60% of pancreatic adenocarcinomas and may be associated with short patient survival.

The tumour suppressor gene *DPC4* (deleted in pancreatic carcinoma, locus 4) resides on chromosome 18q and has homology to a family of proteins called the Mad proteins which play a role in signal transduction for the transforming growth factor (TGF)-β family of cell surface receptors. TGF-β can downregulate the growth of epithelial cells and stimulate differentiation and apoptosis. Therefore, it is reasonable to expect that the loss of DPC4 would promote cell growth. DPC4 is inactivated in approximately half of all pancreatic cancers and, in contrast to *p53* and *p16*, appears to be relatively specific to pancreatic cancer.

Pancreatic intraepithelial neoplasia

It is now believed that invasive pancreatic ductal carcinoma arises from precursor ductal lesions referred to as pancreatic intraepithelial neoplasia (PanIN) and are grouped into three histological grades of increasing degrees of architectural and nuclear atypia (PanIN1a,b, PanIN2, PanIN3). PanINs have led to a better understanding of the sequential molecular changes leading to invasive ductal adenocarcinoma (PDAC) and their study has identified loss of heterozygosity at a number of loci and for alterations in a number of abnormal genes and proteins that are commonly found in pancreatic carcinomas, e.g. K-*ras*, *HER-2/neu*, *p16*, *p21*, *p53*, *DPC4* and *BRCA*. These studies have confirmed that PanINs accumulate genetic changes of increasing atypia, supporting the theory that they are the precursors of PDAC.

S100P has been identified among the genes expressed in PDAC and pancreatic tumour cell lines. S100P is a member of the S100 family of EF-hand, calcium-binding proteins which have a variety of functions and target proteins. S100P is upregulated in intraductal papillary mucinous neoplasms. S100P binds with a protein, S100P–BPR, the interaction being dependent on Ca^{2+} and Mg^{2+} and results in conformational changes in S100P with exposure of its hydrophobic surfaces. S100P–BPR is located at 1p34.3. S100P–BPR is a widely expressed protein and is present in pancreatic islets together with S100P, but neither protein is expressed in normal ductal cells. In contrast, both S100P and S100P–BPR are overexpressed in PanINs and PDAC samples. Furthermore, the overexpression of S100P correlates significantly with increasing grade of PanIN lesions. The similarity in the expression patterns of these two proteins, in healthy pancreas and in PDAC, supports the view that their upregulation is involved in the development of pancreatic cancer. As not all PanINs progress to malignant disease, further studies are needed to identify the genes that determine which high-risk PanINs are likely to progress to invasive PDAC. Alterations in the sequence or expression levels of K-Ras, HER2/neu, p16INK4a, BRCA2, p53, DPC4, cyclin D1, and p21 have previously been described in these precursor lesions. Marked overexpression of S100P in PDAC has been well documented. Its expression in PanIN-1A lesions is low (2%) but increases significantly to 13% in PanIN-1B, 31% in PanIN-2 and 41% in PanIN-3 lesions.

K-Ras activation and overexpression of HER2/neu are the earliest changes in precursor lesions as they are observed in PanIN-1A and PanIN-IB. These changes are followed by an increase in the number of p21-expressing lesions (32% of PanIN-1B are positive for p21) and this is thought to be a direct result of Ras activation or HER2/neu overexpression. S100P expression is one of the genes whose expression is altered very early in the development of PDAC, whereas loss of p16 is observed slightly later (PanIN-2 and -3). Altered p53 expression, cyclin D1 elevation, decreased DPC4, and BRCA2 loss of heterozygosity are later aberrations and occur at a lower frequency. Expression of S100P is twice as prevalent in invasive PDAC (92% of cases) as in PanIN-3 lesions (41% of cases), suggesting that S100P may play an important role in progression from PanINs to invasive PDAC. Thus S100P may prove to be a valuable marker for the prediction of clinically relevant PanINs, i.e. those which are likely to progress to invasive malignancy and thus require intervention.

Pancreatic cancer and diabetes mellitus

There is no doubt that there is an association between pancreatic cancer and diabetes mellitus. Two hypotheses have been put forward. First, diabetes mellitus is an aetiological factor in the development of pancreatic cancer. The evidence for this theory is rather confusing. When recent diabetes was eliminated, some studies found as high as a sixfold risk for pancreatic cancer in diabetic women but not in diabetic men. Several uncertainties result from the fact that there has been no strict delineation of the type of diabetic who is prone to develop pancreatic cancer and no sorting out of the genetic aspects of the disease in the population studied. In addition, there have been varying definitions of diabetes in all reported studies. We therefore do not know if all diabetics or a special subset are at risk from pancreatic cancer. It may well be that diabetics have so many other complications that few actually live long enough to develop pancreatic cancer.

The second hypothesis is that the presence of pancreatic cancer in some way induces glucose intolerance. This is supported by the fact that in many cases the diabetes is diagnosed within 2 years before the cancer is discovered. Thus, there may be two types of diabetes mellitus in pancreatic cancer patients: (1) individuals in whom the hereditary type is present with its possible increased risk of pancreatic cancer and (2) patients in whom the hyperglycaemia is of shorter duration and is a result of the pancreatic cancer. There is a definite suggestion of a bimodality of duration of clinical diabetes mellitus in several series of pancreatic cancer patients (40% of patients with duration of greater than 2 years, and 50% with duration of less than 1 year).

Alcoholism and chronic pancreatitis

Retrospective epidemiological data regarding an association between alcoholism and pancreatic cancer are inconclusive. As with diabetics, alcoholics have so many other problems that pancreatic cancer is one of their lesser worries. The main reason for considering an alcohol–pancreatic cancer association is that pancreatitis (which can be induced by alcohol) has been associated with pancreatic cancer. However, it must be emphasized that 'pancreatitis' may have three different meanings: (1) histological pancreatitis invariably coexists with pancreatic cancer, presumably due to ductal obstruction and/or direct destruction of parenchymatous tissues; (2) the acquired variety of chronic pancreatitis (clinical entity) does not seem to be related to pancreatic cancer; and (3) the hereditary type of chronic pancreatitis seems to have a higher predisposition to pancreatic cancer than the general population.

Cancer of the head of the pancreas

This has to be differentiated from cancer of the ampulla, lower common bile duct and/or duodenum since these latter tumours may present with similar features. The term 'periampullary carcinoma' is often used to denote tumours in this region irrespective of the exact site of origin. Progressive jaundice occurs in over 75% of patients with carcinoma of the head of the pancreas and the incidence of jaundice decreases as the location of the lesion progresses to the left towards the tail of the pancreas. Occasionally, a tumour may invade and compress the third or fourth parts of the duodenum without actually obstructing the common bile duct. Pain is extremely frequent and the classic description of painless jaundice is rarely encountered. Weight loss and anorexia are also common symptoms even in early stages. Nausea, epigastric bloating, change in bowel habits and vomiting are occasionally present. Haematemesis and melaena occasionally occur in late cases as a result of direct invasion of the duodenal or gastric mucosa by tumour or superior mesenteric–splenic vein compression by the tumour. Chills and fever due to cholangitis can occur in longstanding biliary obstruction. A palpable gallbladder (Courvoisier's sign) is noted in only about a quarter of patients with resectable tumours. The liver is usually enlarged on palpation.

Cancer of the body and tail of the pancreas

Pain and weight loss are the two main consistent symptoms. The pain may initially be dull and vague, localized to the epigastrium or to the back, or it may move to either upper quadrant. It may be episodic and related to meals or it may become constant and severe. In late cases, the patient learns to obtain partial relief by flexing the trunk forward. Severe pain invariably indicates extension of tumour into the perineural lymphatics and the posterior parietes. Weight loss is rapid and severe by the time the patient presents to the hospital. Again, haematemesis and melaena may be late features due to mucosal invasion or portal hypertension. Migratory thrombophlebitis (Trousseau's sign) can be present in any patient with advanced cancer, is not specifically indicative of pancreatic carcinoma and, by itself, does not merit diagnostic laparotomy or laparoscopy.

Physical examination in the early stages may reveal surprisingly few abnormal physical signs. In late cases, abdominal masses or liver metastases may be palpable. A rectal shelf may be evident on rectal examination in the rectovesical or rectovaginal pouch (Blumer's shelf), there may be evidence of ascites and distant metastases may be present in the supraclavicular fossa (Troisier's sign).

Delay in diagnosis

Over 90% of patients with pancreatic cancer present in the late stage of their disease at a time when there is no chance of cure and, often, even meaningful palliation cannot be achieved. The factors responsible for late diagnosis are listed below.

- The tumour is asymptomatic in the early stages. There is some evidence that the preclinical phase of pancreatic cancer may be present for months or even years before the tumour 'appears'.
- 'Patient delay' – the early symptoms are often vague and non-specific and the patient tolerates the discomfort.
- 'Physician delay' – the physician often does not have a high index of suspicion and fails to properly 'evaluate' the patient in the face of a vague history and normal physical examination.
- The patient may not have ready and easy access to competent diagnostic centres. Centralization or regionalization of the management of difficult pancreatic problems is long overdue because of the dependence on sophisticated diagnostic and therapeutic methods.

Positive physical signs in a patient with pancreatic cancer often reflect incurability. The diagnosis therefore needs to be made before the appearance of abnormal physical signs. The clinician should always consider the diagnosis of pancreatic cancer in any patient presenting with seemingly genuine recent symptoms, absent physical signs and negative routine radiological investigation. These are the very patients in whom maximum benefit may be gained by applying the more sophisticated investigative techniques.

Evaluation of diagnostic tests

It is no great triumph to diagnose incurable advanced pancreatic cancer. Any particular technique must be assessed on its ability to diagnose potentially curable lesions of the pancreas. A rational sequence of testing is as follows. CT is the best initial test in the evaluation of a patient with suspected pancreatic cancer. Thin section multidetector helical angio-CT scanning through the pancreas with an intravenous bolus injection of contrast can delineate the pancreatic mass, the relationship of the tumour to the superior mesenteric artery and vein and to the coeliac axis, the patency of the portal vein, and the presence of distant metastastic disease.

Until recently, ERCP combined with cytology has been used to differentiate choledocholithiasis from malignant obstruction of the distal common bile duct when a mass is not seen on CT. With the advent of EUS, the rare but real complication of pancreatitis secondary to injection of contrast into the pancreatic duct during ERCP may be avoided. However, EUS requires a dedicated and skilled endoscopist with the appropriate equipment to be successful and may not be available at all medical centres.

It has become obvious that the diagnosis of *advanced* pancreatic cancer can be made after a careful history and routine physical examination. If obvious metastases are present such as seedings in the retrovesical pouch or the pouch of Douglas on rectal or pelvic examination or large left supraclavicular nodes (Troisier's sign) and/or obvious nodular hepatomegaly, careful consideration should be given in avoiding prolonged and unnecessary investigations and even a diagnostic laparotomy. This logic pertains especially to cancer of the body and tail of the pancreas which is rarely, if ever, curable in a symptomatic patient. Percutaneous needle biopsy of any accessible lesion, including the pancreatic mass, can achieve the diagnosis in many cases and the duration of hospitalization can be appreciably shortened. The frail elderly patient with clinically obvious cancer of the body or tail of the pancreas should be spared the mortality and morbidity of a diagnostic laparotomy. Direct percutaneous needle aspiration of the mass with ultrasound or CT scan for guidance should be attempted (Figures 27.20 and 27.21). Another way of obviating laparotomy in these seriously ill people is to perform laparoscopy and direct vision biopsy.

ERCP with stent placement can be performed if the patient is jaundiced, or, if the patient is fit and relatively young, exploration with a view to internal biliary drainage should be performed as described later. The technology of imaging methods is advancing rapidly: 3.0 T MRI, multidetector helical angio-CT with the ability for both image segmentation and 3D volume imaging of the pathology, high-frequency EUS and positron emission tomography (PET) scanning appear to be the

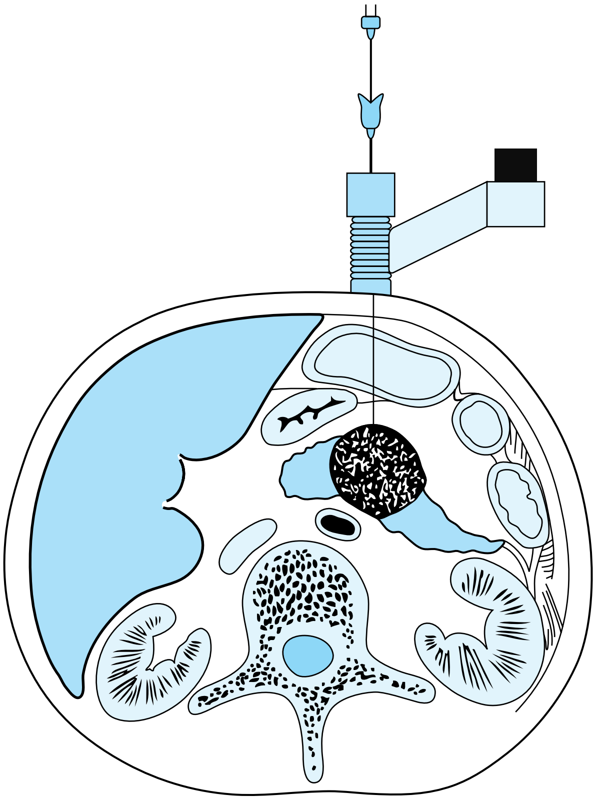


Figure 27.20 Schematic representation of the percutaneous fine needle aspiration technique of a pancreatic mass under ultrasound or CT scan control. The aspirated material is smeared on glass slides and fixed and stained by the Papanicolaou or Giemsa method for microscopic examination.

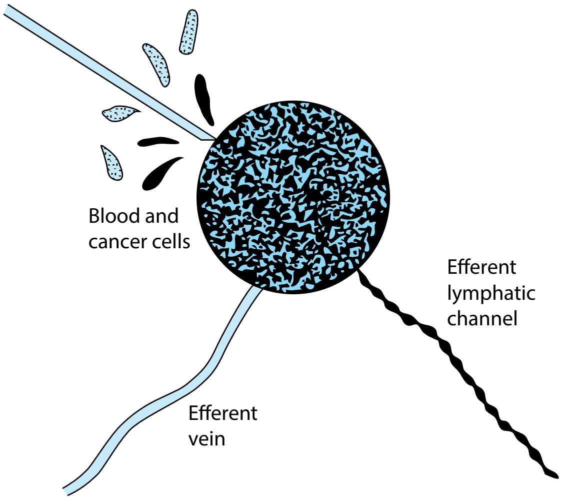


Figure 27.21 Aspiration 'biopsy' and tumour seeding. Seeding of tumour cells occurs experimentally but is probably of no clinical significance in advanced cases.

way of the future, although they are likely to be restricted to specialized regional high-volume centres.

Positron emission tomography scanning

The published evidence suggests that PET imaging is very useful after surgery to detect locoregional recurrence or distant metastases when the follow-up levels of the tumour markers become elevated as PET detects viable metabolically active tumour (as distinct from scarring and other postsurgical benign changes) and which no other imaging modality can. Some surgeons also use PET in the preoperative work-up, especially when the CT or MRI provides inconclusive reports.

PET scan parameters such as maximum tumour standardized uptake value (SUV_m) and the product of tumour SUV and volume [metabolic tumour burden (MTB)] have been shown to correlate with survival. One study on 56 patients with unresectable pancreatic cancer treated with a single fraction of 25 Gy stereotactic body radiotherapy and gemcitibine-based chemotherapy showed that SUV_m and MTB correlated with survival. The subgroup with both lower SUV_m and lower MTB showed a trend towards longer survival with a median of 18.7 months.

Surgical treatment of pancreatic cancer

Emphasis must be placed on preoperative evaluation and adequate preparation of the patient with pancreatic cancer. As mentioned earlier, CT scan, EUS and, in some instances, ERCP and FNA for cytology provide the surgeon with valuable preoperative information and obviate the need for time-consuming manoeuvres on the operating table. It cannot be overemphasized that pancreatic exploration with a view to resection should not be performed by the occasional surgeon or the resident registrar in training or in institutions where there is not all the back-up expertise (endoscopy, radiology, cytology, endocrinology) necessary for the care and management of these difficult problems.

Preoperative preparation

All jaundiced patients must be kept in a good state of nutrition and hydration with supplemental intravenous fluids, elemental diet and multivitamins as deemed necessary. Injudicious use of intravenous contrast for CT scanning may precipitate renal failure, which can also occur postoperatively due to hypovolaemia. A continuous diuresis must be maintained at all times. If the patient is grossly malnourished, a period of parenteral hyperalimentation both before and after operation may be beneficial.

Blood-clotting deficiencies must be corrected. Anaemia is corrected by blood transfusion. Daily injections of vitamin K are administered, preferably for 4–5 days prior to operation. Six units of fresh frozen plasma, six units of platelets and at least six units of blood should be made available. It must be emphasized that pancreaticoduodenectomy can now be safely performed without blood transfusion in many instances.

Cardiopulmonary function should be carefully assessed by pulmonary function tests, chest radiograph and ECG as deemed necessary. Smoking is prohibited. Intensive pulmonary physiotherapy, active mobilization and leg exercises are strongly encouraged preoperatively. The question of prophylactic digitalization and diuretic therapy is considered in individual patients to achieve maximum cardiovascular compensation.

If a patient is critically ill with one or more of the following parameters, (1) a highly elevated serum bilirubin (greater than 200 μm/L (12 mg%); (2) sepsis; (3) hepatorenal failure; (4) severe cardiopulmonary disease which is expected to respond to medical management; or (5) severe malnutrition, a percutaneous transhepatic or endoscopic biliary decompression may be attempted to tide over the patient for 2 or 3 weeks until his/her general condition improves adequately for him/her to be considered for a major pancreatic resection. If the technique of percutaneous biliary drainage or endoscopic stenting is not available, a simple cholecystostomy using ultrasound for guidance may be undertaken under local anaesthesia. However, routine biliary preoperative decompression is *not* recommended as it increases the postoperative mortality and morbidity and should be undertaken only if the patient has septic cholangitis.

Selection of patients for pancreatic resection

Except in unusual circumstances, a major pancreatic resection is inadvisable in (1) elderly people (older than 80 years), (2) frail patients with multiple systemic disorders and (3) those with an estimated life expectancy of less than 3 years. The operation should be reserved for the relatively fit patient under the most favourable circumstances. The surgeon must use clinical judgement in the determination of the relative indications and contraindications for each procedure. The operative mortality should not exceed 5%. A frank discussion must take place between the surgeon, the patient and his or her relatives prior to embarking on a potentially hazardous operation.

Surgical options

For curative surgical treatment of a cancer in the head of the pancreas, four options are available.

1. The Whipple operation (pancreaticoduodenectomy), in which the head and neck of the pancreas together with the duodenum, the distal half of the stomach, lower common bile duct and gallbladder and upper jejunum are removed with as much of the regional lymph nodes as possible.
2. Pylorus-preserving pancreaticoduodenectomy as introduced by Traverso and Longmire in 1978 in an attempt to eliminate the postgastrectomy symptoms seen with antrectomy. This operation technically differs from the standard Whipple procedure only in the preservation of the blood supply to the proximal duodenum.
3. Total pancreatectomy (total pancreaticoduodenectomy) includes, along with the contents of the Whipple operation, excision of the spleen, body and tail of the pancreas, and a more thorough regional lymphadenectomy.
4. Regional pancreatectomy as proposed by Fortner entails extirpation of the transpancreatic portion of the portal vein, the coeliac axis, the superior mesenteric artery and the middle colic vessels, together with structures included in a total pancreatectomy.

Of these four alternatives, the Whipple operation is most commonly employed for tumours of the head of the pancreas. The operation is optimal for malignant tumours that are confined to the duodenum, ampulla of Vater or lower common bile duct. The neck of the gland is divided to the left of the superior mesenteric vein and the body and tail of the pancreas and spleen are left undisturbed. *En bloc* excision of the regional lymph nodes from the porta hepatis, aortocaval and superior mesenteric regions again forms part of the operation. With the Whipple operation, diabetic function can be preserved. Although there is the possibility of an anastomotic leak from the pancreaticojejunostomy, this complication occurs in less than 10% of patients at centres experienced with pancreatic surgery. Also, more effective management of pancreatic anastomotic leakage with hyperalimentation, percutaneous drainage and somatostatin analogue has reduced the magnitude of this problem. Total pancreatectomy should be reserved for situations when there is tumour at the pancreatic margin on serial frozen sections or if the pancreas is not suitable for an anastomosis, especially if the patient is already an insulin-requiring diabetic. A pylorus-preserving Whipple operation is a reasonable alternative but may result in transient gastric stasis.

The concept of extended resection for pancreatic cancer with resections of one or more of the major vessels (regional pancreatectomy) is uniformly attended by an increased morbidity and mortality without a concomitant improvement in cure rate. When such extensive procedures are needed to resect the local tumour, occult metastatic disease is usually present and the disease is incurable. Several authors have advocated a selective approach to venous resection when the lesion has been deemed resectable, the pancreatic neck is divided, and while dissecting the uncinate process from the superior mesenteric vein the tumour is found to be adherent to the posterolateral portion of the vein. The venous segment can be replaced with an internal jugular vein interposition graft. It must be emphasized that resection of the portal–superior mesenteric venous axis is only recommended if it is relatively minor (less than 1 cm in length and less than half of the venous circumference) and it helps in achieving adequate clearance of soft tissue margins.

Operative diagnosis of pancreatic cancer and its differentiation from chronic pancreatitis

The jaundiced patient nowadays is well investigated preoperatively and usually a diagnosis is made prior to laparotomy. A general rule is as follows: hard, non-cystic masses in the head of the pancreas which are associated with obstructive jaundice and dilatation of the common bile duct are usually carcinoma, especially if acute inflammation and/or gallstones are absent. Conversely, hard, non-cystic masses involving a major part of the retroampullary part of the gland and unassociated with jaundice or dilatation of the biliary tree are usually pancreatitis.

It is important to remember that pancreatitis of varying degree invariably coexists with all carcinomas and that patients with gallstones may have a concomitant pancreatic cancer. In doubtful situations, the surgeon must decide whether to try to establish a tissue diagnosis by frozen section histology of biopsies prior to assessing resectability of any pancreatic mass. Every surgeon is influenced by his own philosophy, his experience and expertise, his pathologist's experience and by the clinical situation. Hence, any decision in such a clinical setting can easily lead to controversy when discussed retrospectively or hypothetically. The author's general policy about pancreatic biopsy can be summed up as follows. (1) Since over 75% of all cancers in the head of the pancreas are identified preoperatively by a positive cytology (at ERCP or percutaneous biopsy), it is preferable to assess the resectability of all such masses in the first instance. If conditions are favourable, they are resected without a preliminary biopsy. (2) A suspected cancer of the body and tail of the pancreas is 'biopsied' by a distal pancreatectomy and splenectomy provided that the lesion is localized and resectable. If frozen section histological examination reveals a cancer, the 'excision biopsy' is converted to a total or near-total pancreatectomy with regional lymphadenectomy. (3) All unresectable and/or metastatic tumours of the pancreas are diagnosed before the surgeon leaves the operating room even if the job is time consuming. This takes the matter out of the realm of doubt, an especially important point when a palliative procedure restores the patient to relatively good health for a long period and doubt is raised as to the true diagnosis. A known positive biopsy for adenocarcinoma of the pancreas will then prevent a fruitless second laparotomy. (4) Frozen sections and histological examination of lymph nodes peripheral or adjacent to delineated pancreatic masses are acceptable if unresectability and/or the presence of metastatic disease is documented. However, a positive regional node *per se* is not an absolute criterion of unresectability. Many patients have survived longer than 3 years following pancreatoduodenectomy in the presence of regional lymph node involvement.

This policy is supported by the following arguments.

- Truly representative needle biopsies of the pancreas are often hard to obtain because of sampling error and confusion between tumour and associated pancreatitis (Figure 27.22).
- The establishment of diagnosis by frozen section histology is sometimes time-consuming and traumatic. Factors which influence the biopsy policy of surgeons include personal experience of complications, interpretative histological errors and traditional teaching. Many senior surgeons still regard pancreatic biopsy as inaccurate and dangerous. Pancreatitis, fistula formation, haemorrhage and infection have all been reported following all biopsy techniques. However, it is often difficult to decide whether such complications are directly attributable to the biopsy itself or to the concomitant surgical manoeuvres and manipulations. The consensus of opinion is that all surgeons involved in the practice of pancreatic surgery should be willing and able to perform pancreatic biopsy safely if it is indicated. When this is done, the surgeon should make great efforts to avoid major pancreatic ducts and vessels. A good Kocher manoeuvre between the fingers and thumb of the left hand will guide the biopsy needle through the duodenum into the appropriate suspicious area of the pancreas. The disposable Travenol 'Tru-Cut' needle is very suitable for this purpose. A less traumatic alternative is to employ a long no. 21 needle, pushing it in a similar fashion through the duodenum, and attaching a 10 mL syringe for aspiration in order to provide a smear for cytology. This is gaining in popularity with the more widespread availability of, and co-operation with, skilled cytopathologists (Figure 27.23). Haemorrhage, when encountered, can be controlled by finger pressure. Occasionally, silk mattress sutures are needed for haemostasis of the pancreatic capsule following direct pancreatic puncture.
- Errors in interpretation of frozen section biopsy specimens of the pancreas are sometimes made by the inexperienced histopathologist because some desmoplastic carcinomas closely resemble chronic pancreatitis.

Assessment of resectability

A cancer of the pancreas is considered unresectable if there are distant (liver or peritoneal) metastases, invasion of major vessels (portal vein, hepatic artery, superior mesenteric vessels or coeliac axis) or any extension beyond the area of usual total pancreatectomy specimen. The possible exception is the case of isolated portal vein invasion provided the vein is patent. In these selected cases, portal vein resection with

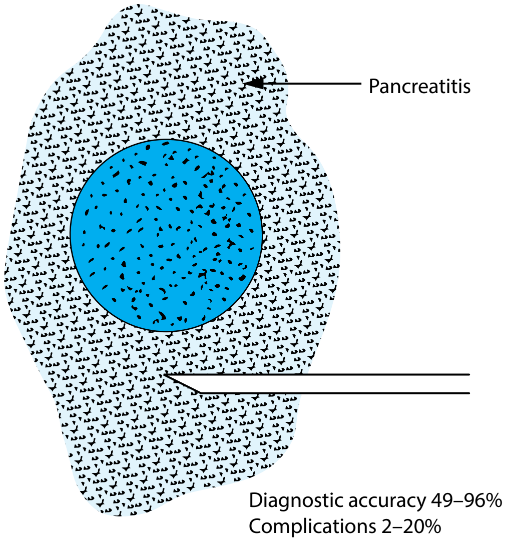


Figure 27.22 Peroperative pancreatic biopsy. A variable degree of pancreatitis surrounds any pancreatic cancer leading to a 'sampling error' with needle biopsies. The number of biopsies taken is limited because of potential morbidity and even mortality and the biopsies may miss the cancer altogether.

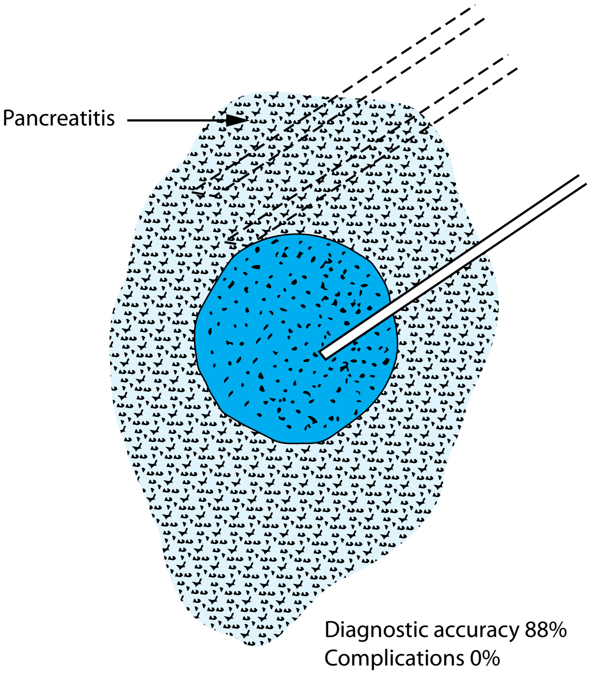


Figure 27.23 Fine-needle aspiration 'biopsy' for cytological examination. Multiple punctures may be performed with minimal risks and increase the chance of obtaining material from the cancer.

interposition venous graft placement has been described. Puckering of the transverse mesocolon *per se* does not always indicate unresectability since the transverse mesocolon and, if necessary, the transverse colon can be excised in a pancreatoduodenectomy specimen if there are no other contraindications to a resection.

Immediate postoperative care and complications

Following a major pancreatic resection, the patient should be transferred to an intensive care unit where experienced nursing care and sophisticated monitoring techniques are readily available. Patients may require respiratory assistance for the first 12–24 hours. For the first 3–4 postoperative days, the patient's blood sugar is checked every 3–4 hours and small doses of regular insulin (2–5 units) are given intravenously as boluses. Alternatively, a continuous intravenous infusion of insulin can be given. It is important to maintain the blood sugar between 75 and 150 mg% (4–8 mmol/L).

Haemorrhage is still the commonest intraoperative and postoperative complication encountered with pancreatoduodenectomy or total pancreatectomy. However, the incidence of this complication has decreased from about 10% to less than 1% in experienced hands. Meticulous preoperative preparation, careful haemostasis, and adequate replacement of blood and clotting factors during the operation are essential. In spite of these precautions, the patient may occasionally continue to bleed at a fairly alarming rate from all raw areas in the abdominal cavity during the first 24 hours. The indications for reoperation are:

- if there is reason to suspect a major bleeding site
- when clot accumulation in the abdomen causes distension and tamponade
- when a consumption coagulopathy is recognized.

In most cases a discrete bleeding point is never found at reoperation. The clots are gently evacuated and the whole abdomen is irrigated prior to closure with drainage. Haemobilia following biliary decompression is not unusual. It invariably stops spontaneously. Whenever haemorrhage is suspected, the patient must be kept normovolaemic by adequate blood and fluid replacement and by maintaining a continuous diuresis. Intermittent doses of diuretics may be given as necessary. Hepatorenal failure is the commonest sequence of events leading to postoperative death in this group of patients.

Other complications which may also be fatal include sepsis, mesenteric thrombosis, uraemia, liver insufficiency, myocardial infarction, cerebrovascular accident, congestive heart failure and pulmonary embolism. Leakage from the biliary enteric anastomosis or from the gastrojejunostomy are largely preventable by careful and proper construction of both anastomoses. Anastomotic leaks from the pancreaticojejunostomy occur in less than 10% of patients at centres experienced with pancreatic surgery. Management of pancreatic anastomotic leakage with hyperalimentation, percutaneous drainage and somatostatin analogue has reduced the magnitude of this problem. Complications that are usually non-fatal include pneumonitis, gastric retention, paralytic ileus, bowel obstruction, wound infection, wound dehiscence, atrial fibrillation, faecal fistula and gastrojejunal fistula.

Recurrence of jaundice after pancreatoduodenectomy for cancer of the pancreas

Recurrence of jaundice and/or cholangitis may be seen after pancreatoduodenectomy and may be due to small bowel obstruction. Nausea and vomiting are usually prominent features. The obstruction may be due to recurrent tumour or simply to adhesions. Laparotomy may be indicated to establish the diagnosis and to relieve the obstruction.

Postoperative chemoradiation for resectable cancer of the pancreas

Data from several randomized trials have shown a significant survival benefit in patients treated with infusional 5-fluorouracil (5FU) and external beam radiation following pancreatectomy. Therefore, in patients who are well enough to tolerate it, adjuvant chemoradiation is recommended. Newer agents such as gemcitabine are increasingly used in combination with radiation as an alternative to 5FU. Indeed the first line agent recommended by the National Institute of Health and Clinical Excellence is gemcitabine; other drugs used are 5FU and capecitabine.

Monitoring of recurrence

A small number of patients with resectable pancreatic cancer have elevated levels of tumour markers such as CA19-9, POA or CEA preoperatively. When this is the case, serial monitoring of either marker may be useful in confirming the completeness of surgical excision and in the detection of recurrent pancreatic cancer.

Replacement of pancreatic exocrine function

Adequate pancreatin tablets (Viokase, Pancrease, Creon) must be taken with each meal. The number of tablets must be increased if steatorrhoea develops. The patient is advised to take a low-fat, high-protein and carbohydrate diet in the form of frequent regular small meals. Patients must take acid-reducing agents (H_2-blockers) half an hour before taking the pancreatic enzymes to prevent acid inactivation.

Factors influencing prognosis after resection for ductal adenocarcinoma of the head of the pancreas

The mortality for major pancreatic resection is between 0% and 5% in specialized centres. Death as a result of operative complications usually occurs within the first 2 months of operation. After 2 months and up to 2 years, death is usually due to metastatic pancreatic cancer although a few individuals can present as late as the end of the third year with metastatic disease. If the patient has survived 3 years, the cause of death is usually unrelated to pancreatic cancer.

Four factors appear to determine survival after pancreatoduodenectomy.

- The site of origin of the cancer. Ampullary and distal bile duct tumours have a better prognosis than pancreatic cancer (Figure 27.24).
- The operative mortality and morbidity of the surgical team. It is well established that the operative mortality, morbidity and eventual survival of patients following this and other major operations are dependent on the experience and expertise available in the institution where the operation is performed (case load and expertise). Hospitals where a large volume of the operation is performed have much better overall results than those where the operation is performed occasionally.
- The stage of the tumour.
- The biological behaviour of the tumour.

Whether these last two factors are independent variables or are different manifestations of the same variable is debatable. It is generally agreed that a uniform staging system for pancreatic cancer is clearly desirable. However, a major problem in staging the disease is that it can only be performed retrospectively after an extensive pancreatoduodenal resection. The following parameters are considered important.

- The size of the tumour: lesions less than 2 cm in diameter have a higher resectability rate and the best overall survival. Lesions greater than 4 cm have the lowest resectability rate and the lowest overall survival. Caution must be expressed here concerning intraoperative decision-making based on size alone. It is not unusual for a small lesion to induce a substantial amount of surrounding pancreatitis and create a larger mass effect which impinges on neighbouring structures.
- Positive lymph nodes are an important negative prognostic factor. Hence, it is important to perform an adequate lymphadenectomy *en bloc* with the pancreatoduodenectomy specimen to ensure appropriate staging by the pathologists. How extensive the lymph node dissection and excision of surrounding soft tissues should be is debatable.
- Extrapancreatic invasion adversely affects survival. Since microscopic peripancreatic invasion appears to be important, the extent of the peripancreatic excision leading to skeletonization of the major vessels as advocated by several Japanese surgeons makes sense in order to provide a microscopically curative dissection. Whether a pylorus-preserving pancreatoduodenal resection can be achieved at the same time is debatable.
- Major vascular involvement: this is also an adverse prognostic factor. In most instances, this has the connotation of unresectability if the hepatic artery, coeliac axis or superior mesenteric artery is invaded. However, a short segment of portal vein–superior mesenteric vein axis is not necessarily incompatible with long survival, provided that *en bloc* excision of the adherent venous segment helps achieve a microscopically curative surgical resection.
- Histological grade of the tumour: poorly differentiated tumours are associated with lower survival than well-differentiated ones. Using absorption cytometry to measure DNA content of cells, patients with diploid cancers have a 50% 5 year survival compared with no 5 year survival among those patients with aneuploid cells.
- The amount of blood transfusion in the perioperative period. Just like in the case of colon cancer, this factor has been proposed as an independent variable affecting survival. Patients who underwent pancreatoduodenectomy with 2 units or less of blood transfusion have a median survival of 24.7 months, whereas those who required 3 units or more in the perioperative period survive only a median of 10.2 months.
- Perineural involvement. Extensive microscopic involvement of perineural lymphatics is usually associated with a poor prognosis, but such tumours are, by and large, extensive in terms of both peripancreatic involvement and nodal metastases.
- Sex of the patient: women appear to live longer than men after pancreatoduodenal resection for pancreatic cancer.

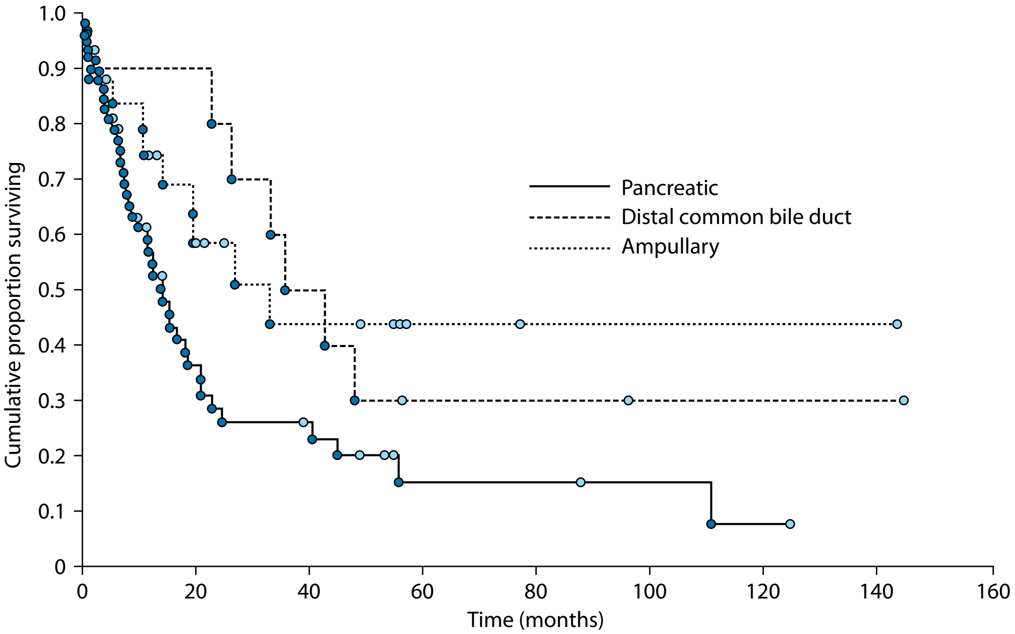


Figure 27.24 Different survival rates after resection of pancreatic cancers, cancers of the distal common bile duct and periampullary cancers.

Palliation of pancreatic cancer

The surgeon can palliate incurable pancreatic cancer in several ways.

- Relief of jaundice, pruritus and impending cholangitis: the biliary tract decompression can be done either by cholecystojejunostomy or by hepaticojejunostomy (each with a diverting enteroenterostomy), depending on whether the cystic duct is widely patent and is in full communication with the biliary tree proximal to the obstructing cancer.
- Relief of duodenal obstruction. If the patient lives for more than a few months, duodenal obstruction invariably occurs. It is therefore advisable to perform a gastrojejunostomy at the primary operation. More recently endoscopically inserted wall stents are used for established duodenal obstruction.
- Relief of pain. The coeliac plexus can be infiltrated with 50 mL of 50% alcohol or with 20 mL of 6% phenol. This may be helpful in patients with cancer of the body of pancreas when the pain is a prominent feature. Cordotomy, extensive sympathectomy and stereotactic thalamotomy have all been tried with minimal or no objective response. More recently, thoracoscopic splanchnicectomy has been shown to achieve substantial pain relief in inoperable pancreatic cancer.

All locally unresectable or metastatic masses must be biopsied until a definite histological diagnosis is made on frozen section histology. The main tumour mass must be outlined with silver clips to provide a possible radiation port. Postoperative chemoradiation for locally advanced, unresectable tumours is indicated. If liver metastasis or carcinomatosis is found at laparotomy, single agent chemotherapy with gemcitabine may provide palliation and improve the patient's quality of life.

When a patient is unfit for surgery, or refuses operation, an alternative method of palliating the obstructive jaundice is by endoscopic sphincterotomy and placement of biliary stent. This approach does not relieve any duodenal obstruction which may be present. If the patient survives for more than a few months, recurrent cholangitis associated with stent blockage is a problem that necessitates regular endoscopic removal and replacement of the stent. Newer self-expandable metallic devices such as the wall stents have improved patency rates, and are nowadays used in preference to plastic stents. Percutaneous transhepatic placement of an internal expandable metal stent inserted by the interventional radiologists is an alternative option for palliation of the jaundiced patient with malignant biliary obstruction.

Intraduct papillary mucinous neoplasms of the pancreas

Mucinous tumours of the pancreas include serous cystic lesions, mucinous cystic neoplasms, solid pseudopapillary neoplasms, cystic islet cell tumours and intraductal papillary mucinous neoplasms of the pancreas (IPMNs). Although sometimes referred to as mucinous duct ectasia or intraductal papillary mucinous tumours, the term IPMNs or IPMTs is generally used. These tumours are composed of mucin-producing columnar cells with papillary proliferation and cyst formation and may involve the main pancreatic duct, the branch ducts, or both. First described in Japan in 1982 in a small series of patients with dilated main pancreatic ducts, patulous ampullary orifices and mucus secretion from the pancreatic duct, they were thought to be rare tumours, but their incidence has increased markedly during the last decade, mainly due to increased diagnosis by advances in medical imaging. The distinction between IPMN and mucinous cystic neoplasm (MCN) was not established before 1999, and it is now thought that many lesions previously classified as MCNs were in fact IPMNs.

Pathology

IPMNs are primarily intraduct tumours with a papillomatous growth pattern, which is associated with excessive mucin secretion and results in progressive ductal dilatation or cyst formation. IPMNs occur in four forms: (1) segmental involvement of the main pancreatic duct, (2) diffuse involvement of the main pancreatic duct, (3) macrocystic involvement of a branch duct and (4) microcystic involvement of a branch duct. The correct diagnosis, previously only possible with ERCP, can now be made with CT and MRI-MRCP.

Pathologically, IPMNs can be divided into the three categories according to the criteria established by the World Health Organization. These are:

- IPMN adenoma
- IPMN borderline
- intraductal papillary mucinous carcinoma.

Intraductal papillary mucinous carcinoma is further categorized into non-invasive (carcinoma *in situ*) and invasive carcinoma. The prevalence of malignant lesions associated with IPMN, including carcinoma *in situ* and invasive carcinoma, is variously reported to range widely between 30% and 88%. IPMNs represent a spectrum of neoplasms with varying aggressive behaviour from small benign adenomas to aggressive invasive cancer. Although IPMNs carry a relatively more favourable prognosis than ductal adenocarcinoma of the pancreas, these lesions are established precursors to invasive adenocarcinoma of the pancreas. The preoperative determination of the benign or malignant type of IPMN in the individual patient dictates the surgical treatment and determines the prognosis. The features indicative of malignancy include:

- marked dilatation of the main pancreatic duct (in some but not all studies)
- main duct type or combined type lesions
- large mural nodules
- solid mass
- diffuse or multifocal involvement
- large tumours of the branch duct type (in some but not all studies)
- widely opened orifice of the papilla of Vater on ERCP.

Clinical features

The true incidence of IPMNs is not known. However, it is thought that IPMNs account for 1–3% of neoplasms of the exocrine pancreas and 20–50% of cystic pancreatic neoplasms. These tumours are usually diagnosed in the elderly (fifth to sixth decade) with a male-to-female ratio for main duct IPMNs of 1.1 to 3:1, and 0.7 to 1.8 for branch duct IPMN. However, the sex ratio varies geographically, with a male predominance in Japan and Korea and a more even distribution in Western countries.

Clinical features suggestive of malignancy include a recent onset or worsening of diabetes, steatorrhoea and jaundice. In contrast, patients with benign IPMNs present more commonly with abdominal pain and a longer duration of the symptoms.

Surgical treatment

Case selection for surgical treatment vs surveillance depends on the assessment of the lesion by ERCP and other imaging tests, and the patient's age and fitness for surgery. Features indicative of malignancy (outlined in the previous section) are indications for surgical resection in all fit patients, irrespective of age. Expectant policy with surveillance is indicated in asymptomatic patients with branch IPMNs, absence of any features (clinical or radiological) of malignancy and those at high operative risk because of comorbid disease.

As large duct IPMN affects the head of the pancreas and the uncinate process (60%) and less commonly the entire gland, the recommended treatment is pancreaticoduodenectomy for proximal disease (with frozen section clear resection margins) and total pancreatectomy. All patients with dilatation of the entire ductal system of the gland require total pancreatectomy. IPMNs involving only a segment of the pancreas (branch duct disease) present problems in adequate complete clearance. In these cases, intraoperative frozen section histology for establishing tumour-free resection margins is mandatory. The surgeon must continue the resection until negative margins are obtained. Even so, negative resection margins carry a reported local recurrence rate of from 0% to 25% depending on the length of follow-up.

Lesions of the endocrine pancreas

Pancreatic islet cells are components of the gastroenteropancreatic part of the diffuse neuroendocrine system. Cells belonging to this system are commonly referred to as APUD cells because they share the following cytochemical characteristics: a high amine content (A); the capacity for amine precursor uptake (PU); and the property of decarboxylation (D) of these precursors to form amines. Tumours arising from the APUD series of cells are called APUDomas. Although Pearse originally suggested that all APUD cells were probably derived from neural crest cells, it is now generally recognized that the gastroenteropancreatic APUD cells (including pancreatic islet cells) probably arise from endoderm (dedifferentiation theory). Regardless of their origin, the APUD cells share similarities in structure, properties and potential. All have characteristic histological, histochemical, immunocytochemical and electron microscopic appearances and all contain the enzyme neurone-specific enolase that is the universal marker for such cells and their hyperplastic and neoplastic lesions.

APUD cells are capable of synthesizing and secreting a great variety of peptides which exert regulatory effects by four main modes of action:

- endocrine, i.e. involving secretion into the circulation to affect distant target sites
- paracrine, i.e. involving local secretion to act on adjacent cells
- neurocrine, i.e. involving secretion at neuronal synapses to act as a neurotransmitter
- neuroendocrine, i.e. involving release of a peptide product of the neurone into the circulation to act on other tissues.

Clinical syndromes may develop as a result of either inadequate or excessive production/release of the potent chemical messengers (e.g. inadequate insulin causing diabetes mellitus and excessive insulin leading to hypoglycaemia). The development of radioimmunoassays for a number of the gastroenteropancreatic peptide hormones has led to the understanding that hyperplastic as well as neoplastic pancreatic islet cells are capable of producing recognizable syndromes of hormone excess. Pancreatic tumours of the islet cell type may in fact secrete two or more identifiable peptide hormones although the threshold for the appearance of their respective clinical symptoms varies greatly and elevated levels of one hormone may be compensated by regulatory hypersecretion of other hormones. Endocrine tumours of the pancreas can be referred to as entopic if they produce hormones normally secreted by the pancreas (e.g. insulinoma, glucagonoma) and ectopic if they produce non-pancreatic hormones (e.g. gastrinoma, VIPoma). There is little correlation between tumour size and plasma hormone concentration or severity of clinical symptoms. Some 10–25% of patients harbouring pancreatic APUD tumours will have the multiple endocrine neoplasia type 1 (MEN-1) syndrome. The clinical syndromes associated with overproduction of identified pancreatic islet cell peptides are shown in Table 27.6.

The overwhelming majority and perhaps all islet cells additionally cosynthesize and cosecrete the protein chromogranin along with their peptide products. Elevated plasma levels of chromogranin or neurone-specific enolase are useful markers for pancreatic endocrine tumours and the MEN syndromes even in the absence of clinical symptoms or demonstrable hormonal excess. Immunohistochemical staining for these proteins confirms the endocrine nature of otherwise obscure pancreatic tumours.

With the exception of insulinoma, most endocrine pancreatic tumours are frankly malignant or at least have a high potential for metastatic spread. Compared with their non-endocrine pancreatic counterparts, the endocrine cancers are relatively slow growing and many apparently metastasize only to regional lymph nodes. This characteristic allows curative surgical therapy in a sizeable proportion of patients.

Historically, the symptoms, morbidity and mortality of the pancreatic endocrine tumour have been due primarily to hormonal hypersecretion but earlier detection and more effective treatment now can often minimize symptoms and forestall death until the advanced stages of malignant spread.

Overview of pancreatic endocrine tumours

Pancreatic endocrine tumours comprise a heterogeneous group of tumours of varying histological differentiation which are broadly divisible into (1) functioning and (2) non-functioning types. With *functioning tumours* the clinical symptom manifestations are the result of persistent autonomous hypersecretion of peptide/hormone(s). In contrast, *non-functioning tumours*, although capable of secreting various hormones (PP, human chorionic gonadotrophin, calcitonin, neurotensin, etc.), do not usually produce specific clinical symptoms resulting directly from overproduction of hormone(s). In this respect they are considered as *non-functioning* and do not produce specific recognizable syndromes. Some functioning tumours are rare and may also originate in non-pancreatic tissues (extrapancreatic), e.g. VIPomas, somatostatinoma, GRFoma, etc.

Epidemiology of pancreatic endocrine tumours and rare functioning tumours

The incidence of clinically detected pancreatic endocrine tumours is reported as 4–12 cases per million individuals. The commonest functioning pancreatic endocrine tumours are insulinomas (17%), followed by gastrinoma (15%). Collectively the others are classed as rare functioning tumours (RFTs) and include VIPoma (2%), glucagonoma (1%), carcinoid (1%) and somatostatinoma (1%). The majority of RFTs are well-differentiated and most are malignant (WHO group 2) and have usually metastasized to the liver by the time of diagnosis. The 5 year survival for localized disease is however good (variously reported at 60–100%). The survival drops to 40% in patients with regional disease and to 30% in patients with distant metastases.

RFTs which represent less than 10% of all pancreatic endocrine tumours can occur at any age but their average age at diagnosis is 50–55 years. They have an equal sex distribution. Some 25% of patients with pancreatic endocrine tumours have MEN-1 and these patients may develop multiple tumours either synchronously or subsequently (metachronous lesions). In contrast to the more common pancreatic endocrine tumours, the incidence of MEN-1 in patients with RFTs is much lower, i.e. about 2% for VIPomas and glucagonomas.

Detection and localization

The standard imaging procedures for pancreatic endocrine tumours and RFTs include EUS, contrast-enhanced helical CT or MRI of the abdomen in combination with somatostatin receptor scintigraphy (SRS). EUS is of proven efficacy in detecting most pancreatic endocrine tumours and can be combined with EUS-FNA, but is not generally available. SRS is considered a routine investigation for both primary tumours and metastases and should always be performed prior to surgery. Gallium-labelled somatostatin analogue PET can detect small tumours. Recently, reports of PET with 5-HTP or ^{18}F-DOPA have shown promising results in the detection of small well-differentiated tumours.

The laboratory tests for the detection of pancreatic endocrine tumours and RFTs include both specific [insulin, VIP, glucagon, somatostatin, GRF, adrenocorticotrophin hormone (ACTH)] and general (chromogranin A and PP) markers.

Pathology and genetics

Pathological diagnosis is obtained by biopsy and is performed using both haematoxylin and eosin staining and immunohistochemical staining with chromogranin and synaptophysin. The biological aggressiveness of the tumour is determined by the mitotic index obtained by counting 10 high-power fields and by calculation of the Ki-67 index. Genetic testing is indicated in cases of suspected familial predisposition to MEN-1 and in patients with other associated endocrinopathies, e.g. elevated serum calcium (hyperparathyroidism) or parathyroid hormone (prolactinoma).

Surgical treatment

In all patients with functioning tumours, specific measures to avoid hormonal crisis during surgery are essential. These include perioperative somatostatin analogue therapy and special anaesthetic measures. Curative surgery should be attempted even in the presence of localized metastatic disease to the liver. The primary pancreatic tumour is resected but the operation varies from pancreaticoduodenectomy (for large proximal tumours), enucleation (for small tumours of the head) and distal pancreatectomy with or without splenic preservation (for tumours of the distal pancreas). If malignancy is suspected, it should be confirmed by frozen section and the tumour resected with complete regional lymph node clearance. Complete resection of liver metastases should always be considered for both functioning and non-functioning tumours. Intraoperative contact ultrasound is helpful both in the detection and for guiding the anatomical resection of liver metastases. Hepatic resection may be done at the time of excision of the primary or subsequently depending on expertise and extent of hepatic resection necessary.

Palliative surgery (debulking of primary or metastases) may be indicated following case conference with the oncologist, as extensive debulking can control symptoms related to hormonal hypersecretion and improves the efficacy of chemotherapy. Bilateral adrenalectomy is indicated in patients with excess ACTH secretion resulting in Cushing syndrome. Liver transplantation may be indicated for a small subset of patients with diffuse liver disease without extrahepatic metastases in whom life-threatening hormonal symptoms persist despite medical therapy.

Embolization and other methods of local ablation

Selective embolization alone or in combination with intra-arterial chemotherapy (chemoembolization with streptozotocin + doxorubicin or mitomycin C) is an established procedure which is effective in controlling symptoms and retarding progression of tumour and yields good control of symptoms and partial tumour response in 50%. Chemoembolization carries a low mortality rate but is accompanied by significant morbidity due to hepatic or renal failure. In addition, patients often develop the postembolization syndrome consisting of prolonged fever, right upper quadrant pain, nausea, elevation of liver enzymes and a decrease in albumin and prothrombin time. The syndrome is treated with adequate analgesia and hydration together with infusion of somatostatin analogues.

Other local ablative methods, which may be used alone or in combination with surgery include radiofrequency ablation, cryotherapy and laser thermal ablation but these are only effective for limited disease with deposits not exceeding 4–5 cm in diameter.

Medical therapy

Both somatostatin analogues and interferon are effective in the control of symptoms caused by functioning PETs, and this also includes RFTs. Thus somatostatin therapy results in dramatic improvement in 80–90% of patients with VIPoma (control of the diarrhoea) and in patients with glucagonoma (subsidence/minimization of rash) and 60–80% exhibit a reduction in VIP and glucagon levels. However, symptomatic relief is not always accompanied by a reduction in circulating hormone levels, indicating that somatostatin analogues have direct effects on the peripheral target organ. Recurrence of symptoms is encountered frequently. It may respond to an increase in dose of somatostatin analogues but this is usually a temporary measure. Although the antitumour efficacy of somatostatin analogues is low with objective tumour responses of >10%, disease stabilization of up to 40% has been reported especially in subgroups of patients with slowly progressive well-differentiated tumours expressing sst2 receptor subtype (i.e. with a positive SRS). Somatostatin analogue therapy is initiated with short-acting substance (octreotide 100 μg subcutaneously two or three times daily) for 1–2 days with titration according to clinical response. Treatment is then changed to slow-release lanreotide SR i.m., lanreotide autogel s.c. or sandostatin–LAR (long-acting release) i.m. (every 4 weeks). Interferon is indicated in metastatic slow-proliferating tumours and can be effective in VIPomas not responding to somatostatin analogues.

Systemic chemotherapy is indicated in patients with metastatic and progressive disease using combinations of streptozotocin and 5FU and/or doxorubicin with objective response rates averaging 35%. Peptide receptor radionuclide therapy (PRRT) was introduced with the development of chelators suitable for radiometal labelling allowing for coupling of modified somatostatin analogues with trivalent metal ions (indium, gallium, yttrium, lutetium, etc.). The efficacy of PRRT in the treatment of advanced PETs with positive SRS has been demonstrated, although the reported data for RTTs are limited.

Follow-up

The follow-up of PETs and RFTs should include clinical, biological and imaging tests at regular intervals with most patients being assessed at intervals of 3–6 months.

Hyperinsulinism

Hyperinsulinism in its primary form embraces several different varieties of pancreatic islet cell disease which include -cell hyperplasia/microadenomatosis and -cell neoplasia (insulinoma). These conditions manifest as symptomatic hypoglycaemia. Insulinomas are the most common of pancreatic APUDomas (75% of symptomatic cases) and the most frequent cause of organic primary hyperinsulinism. In the adult, approximately 80% of insulinomas are benign solitary tumours. There is an even distribution of tumours in the head, body and tail of the pancreas. Multiple tumours are usually present in patients with MEN-1 syndrome and are found in about 10% of cases. -cell carcinoma occurs in 5–10% and is characterized by local invasion and metastatic spread to regional lymph nodes and liver. Primary hyperinsulinism is rare in infants and children but, when encountered, a form of -cell hyperplasia (nesidioblastosis) is seen much more frequently than neoplasia. In contrast, microadenomatosis or islet cell hyperplasia is only very rarely found in adults.

Clinical features and diagnosis of hyperinsulinism

Hypoglycaemia induces a constellation of symptoms reflecting activation of the autonomic nervous system and release of epinephrine (adrenaline) together with cerebral dysfunction related to insufficient glucose oxidation to meet energy needs. The symptoms of adrenergic hyperactivity are more apt to occur with rapid falls of plasma glucose and include weakness, sweating, hunger, palpitations and tremulousness. Neuroglycopenia manifests as headache, visual disturbance, dizziness and confusion and may progress to abnormal behaviour, seizures and coma. Hypoglycaemic episodes are often misinterpreted as suggesting brain tumour, epilepsy, alcoholism or drug abuse, psychosis or even hysteria. Delays in diagnosis and treatment of hypoglycaemia are common and contribute to the morbidity and mortality of the condition. The most important clue to early correct diagnosis is the relationship of the symptoms to periods of food deprivation or physical exercise and the relief of symptoms following food ingestion. In cases where diagnosis is long delayed, patients often develop obesity from increased carbohydrate intake as a behavioural adaptation to repeated episodes of symptomatic hypoglycaemia. Thus, it is not surprising that the diagnosis of hyperinsulism is often delayed for several years following the onset of symptoms. Patients are often ‘shunted’ from neurologist to cardiologist to psychiatrist.

Differential diagnosis of hypoglycaemia

Hypoglycaemia may occur in the fasting state or may be postprandial (reactive) in nature. In the latter condition, low plasma glucose concentrations occur only in response to meals. In fasting hypoglycaemia, a period of hours to a few days is required to precipitate hypoglycaemia. While patients with fasting hypoglycaemia (particularly insulinomas) may also exhibit a reactive component, patients with reactive hypoglycaemia never have symptoms when food is withdrawn. Fasting hypoglycaemia usually indicates a specific underlying disease process while symptoms suggestive of postprandial hypoglycaemia are often found in the absence of an identifiable organic lesion.

Postprandial (reactive) hypoglycaemia

The more common causes of postprandial hypoglycaemia are:

- alimentary hyperinsulinism
- hereditary fructose intolerance
- galactosaemia
- leucine sensitivity
- idiopathic
  - true hypoglycaemia
  - non-hypoglycaemia.

Alimentary hypoglycaemia is the most common type seen clinically and is usually found in patients who have undergone gastrectomy, pyloroplasty, gastrojejunostomy or, rarely, proximal gastric vagotomy. Symptoms occur within a few hours postprandially and are particularly prominent after meals of high carbohydrate content in the form of mono- and disaccharides. Although the exact pathophysiological mechanisms remain to be defined, it is clear that rapid gastric emptying of simple sugars by the postoperative stomach with brisk absorption of glucose and excessive insulin release are of central importance. The attacks can be provoked in affected individuals by oral ingestion of 100 g of glucose in water. A rapid abnormal rise in plasma glucose together with a parallel, and often exaggerated, plasma insulin response occur following glucose challenge. Hypoglycaemic symptoms appear within 1–2 hours as the insulin response and/or effect exceeds the requirement of euglycaemia. True alimentary hypoglycaemia may occur in the absence of gastrointestinal surgery but is rare.

Reactive hypoglycaemia is often misused as a diagnostic label in patients suffering from anxiety states rather than true idiopathic reactive hypoglycaemia. Although some of these individuals manifest a very mild and asymptomatic depression in plasma glucose during the 5 hour glucose tolerance test, hypoglycaemia cannot be documented after normal meals containing less than 100 g of rapidly absorbable carbohydrate. This is in contrast to the occasional case of true idiopathic reactive hypoglycaemia where spontaneous symptomatic episodes are reproducible and accompanied by demonstrably low plasma glucose levels. Most patients without true hypoglycaemia have postprandial adrenergic discharge as a result of underlying anxiety and stress. The epinephrine (adrenaline)-mediated symptoms suggest hypoglycaemia but occur in its absence and are presumably psychogenic in origin.

Fasting hypoglycaemia

The major causes of fasting hypoglycaemia are:

- conditions primarily due to underproduction of glucose
  - hormone deficiencies: hypopituitarism, adrenal insufficiency, catecholamine deficiency, glucagon deficiency
  - enzyme defects: glucose-6-phosphatase, liver phosphorylase, pyruvate carboxlase, PEP (phosphoenolpyruvate)-carboxykinase, fructose-1, 6-diphosphatase, glycogen synthetase
  - substrate deficiency: ketotic hypoglycaemia of infancy, severe malnutrition, muscle wasting, late pregnancy
  - acquired liver disease: hepatic congestion, severe hepatitis, cirrhosis
  - drugs: alcohol, propranolol, salicylates
- conditions primarily due to overutilization of glucose
  - hyperinsulinism: insulinoma, exogenous insulin, sulfonylureas, immune disease with insulin antibodies
  - appropriate insulin levels: extrapancreatic tumours, cachexia with fat depletion, carnitine deficiency, carnitine acyltransferase deficiency.

In this condition one or both of the following mechanisms may be operative: (1) hepatic glucose production is not adequate to meet ordinary tissue demands; (2) peripheral glucose utilization is increased to such a degree that maximal hepatic production is insufficient to match glucose egress from the plasma component. Since hepatic glucose output in normal fasting man is between 100 and 200 g per day, a requirement for greater than 200 g of intravenous glucose over a 24 hour period to prevent hypoglycaemia can be taken as evidence for overutilization of glucose.

From the practical standpoint, fasting hypoglycaemia in an otherwise healthy individual is always due to hyperinsulinism that is attributable to an insulinoma in the adult or islet cell hyperplasia in the neonate or infant. Although patients with hyperinsulinism classically describe or manifest symptoms under fasting conditions in the early morning hours before breakfast or in the late afternoon following exertion, attacks may be highly unpredictable and distributed randomly throughout the day. While it is obvious that the development of fasting hypoglycaemia in insulinoma patients is due to excessive insulin secretion, the hypoglycaemia may result from insulin-mediated suppression of hepatic glucose production as well as augmentation of glucose utilization. Most patients learn quickly that symptoms can be relieved by intake of food or sweetened drinks. Accordingly, a proportion of patients gain substantial amounts of weight.

In 1935, Whipple and Franz reviewed 35 cases of insulinoma and enunciated the primary diagnostic criteria which became known as Whipple's triad:

- hypoglycaemic symptoms are produced by fasting
- hypoglycaemia is documented during symptomatic episodes (blood glucose level below 50 mg/dL)
- symptoms are relieved by glucose intake.

While the presence of Whipple's triad strongly suggests the presence of an insulinoma, differentiation from other causes of fasting hypoglycaemia is crucial. Factitious hypoglycaemia due to surreptitious self-administration of insulin must always be considered in cases posing diagnostic difficulty. Currently, the diagnosis of insulinoma is based upon three elements:

- recognition of the probable nature of the patient's symptoms
- presence of Whipple's triad
- demonstration that the plasma insulin concentration is inappropriately high for the existing level of plasma glucose.

Thus, it is not the absolute level of insulin but its concentration relative to the plasma glucose that is diagnostic. Although absolute elevation of the insulin level is present in many insulinoma patients, rapid degradation of insulin by the liver is probably responsible for the normal absolute levels seen in others with functioning islet cell tumours. For this reason, the ratio of plasma immunoreactive insulin to plasma glucose is considered of greater diagnostic accuracy than absolute levels of insulin and glucose. An insulin (μg/mL) to glucose (mg/dL) ratio of greater than 0.3 indicates insulinoma. It is therefore essential in the investigation of suspected or documented hypoglycaemia to measure simultaneous insulin and glucose levels from the same plasma sample obtained at the time of hypoglycaemia.

Virtually all insulinoma patients will develop symptomatic hypoglycaemia during a diagnostic 72 hour fast. About 90% will manifest symptoms within 48 hours, 80% within 24 hours and 40% within 2 hours of fasting. The plasma glucose level at the time of symptoms is almost invariably less than 40 mg/dL (2.2 mmol/L). In normal individuals the level of immunoreactive insulin during fasting is very low to almost undetectable. At the time of fasting hypoglycaemia, almost all insulinoma patients have basal insulin levels greater than 5 μg/mL. Other causes of fasting hypoglycaemia such as fibrosarcoma and other non-pancreatic tumours, glucocorticoid deficiency or diffuse liver disease may exhibit a positive Whipple's triad but in none will the associated immunoreactive insulin level be increased.

Plasma proinsulin levels are also helpful in the diagnosis of insulinoma. Proinsulin is the single chain precursor of insulin and is normally present to an extent of 20% or less of the total immunoreactive insulin. Under ordinary circumstances, proinsulin is split into C-peptide (connecting chain) and insulin prior to storage in β-cell granules, with only a small percentage being secreted into the portal venous circulation (Figure 27.25).

Because insulinoma tumour cells are usually less differentiated than normal β-cells, they secrete more proinsulin. Simple determination of proinsulin in overnight fasted plasma provides good separation of patients with islet cell tumour from normal individuals in 90% of cases. Proinsulin can be expressed as the absolute level or as a percentage relative to total insulin concentration in plasma. Occasionally, some well-differentiated insulinomas may have normal proinsulin secretion. Greatly elevated values (greater than 50% of total immunoreactive insulin) are often associated with malignant tumours.

Provocative tests in the diagnosis of insulinoma

While no other test is as simple, safe and accurate as prolonged fasting with measurement of glucose and insulin, on rare occasions provocative testing may be needed. A variety of such tests have been advocated and all can be misleading and some potentially dangerous. The glucose tolerance test and the leucine infusion test are mentioned only to be condemned because of their inaccuracy. None of the others is diagnostic in more than 70% of cases. Their further disadvantage is the provocation of occasionally severe hypoglycaemic reactions resulting from release of substantial amounts of insulin from the tumour.

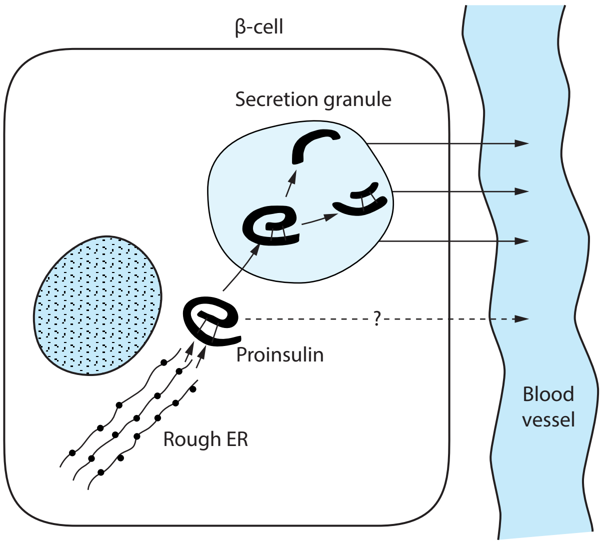


Figure 27.25 Conversion of proinsulin to insulin and C-peptide within the pancreatic β-cell. Equimolar amounts of insulin and C-peptide are liberated during exocytosis. ER signifies endoplasmic reticulum. In normal circumstances only small amounts of the precursor, proinsulin, are released into the bloodstream, while insulinomas release proinsulin into the circulation in larger amounts.

The calcium infusion test has been used as a provocative test for a number of APUDomas, including gastrinoma, medullary thyroid carcinoma and carcinoid tumours as well as insulinomas. Patients with insulinomas release insulin and proinsulin with resultant hypoglycaemia after calcium infusion. In order to avert the hypoglycaemic attack attending calcium infusion-stimulated insulin release, a combined glucose–calcium infusion has been devised in which insulin release after glucose alone and after glucose plus calcium is compared. This, like all stimulation tests, suffers from a relatively low diagnostic accuracy and is indicated only when diagnostic difficulties arise.

The tolbutamide and glucagon tolerance tests are performed by intravenous administration of the respective drug followed by plasma collections for glucose and insulin determinations over a 1 hour period. In the tolbutamide test, a plasma insulin level of 195 μg/mL or greater is considered diagnostic. The critical insulin value in the glucagon test is 160 μg/mL.

A useful suppression test involves infusion of fish insulin to produce hypoglycaemia. While normal subjects respond with suppression of endogenous insulin secretion, patients with insulinoma fail to suppress because of the autonomous nature of hormone release by the tumour cells. Porcine insulin may be used instead of fish insulin but C-peptide must then be measured as a marker of endogenous insulin release. While the non-suppressibility of endocrine tumours is often interpreted as a sign of malignancy, caution in this regard should be exercised with respect to insulinomas where a different degree of suppressibility reflects a different degree of functional dedifferentiation but not necessarily malignancy.

Newer tests for insulinoma which may have good diagnostic accuracy without danger of hypoglycaemia are the sequential suppression tests using somatostatin followed by diazoxide with measurement of insulin levels and glucose consumption and the computer-controlled glucose infusion system which measures the glucose infusion rate required to maintain plasma glucose constant at 80 mg/dL (4.4 μmol/L).

Factitious hypoglycaemia must always be considered and this is especially true in individuals having access to insulin or oral hypoglycaemic agents. Concomitant measurement of plasma levels of C-peptide is critical in such circumstances as insulin and C-peptide are secreted in equimolar amounts by both the normal β-cells and the insulinoma cells. The C-peptide level thus serves as a direct marker of endogenous insulin release. Since all of the diagnostic tests presented can be positive following administration of exogenous insulin, the finding of an inappropriately depressed level of C-peptide can readily identify the surreptitious insulin user.

Preoperative tumour localization

There are some who doubt the value of preoperative imaging of insulinomas as the overall accuracy of preoperative localization tests is low especially for small tumours (<1.0 cm) and averages

30–35% overall including ultrasonography, CT and MRI. Invasive tests [angiography, transhepatic portal venous sampling (THPVS), EUS] can localize lesions in >70%. In contrast, intraoperative inspection and palpation gives superior localization (90%) and this is improved by contact intraoperative ultrasound (93–95%), especially in insulinomas not detected by preoperative imaging (occult insulinomas). Most insulinomas are small benign adenomas with over 75% being less than 1.5 cm in diameter.

The counter argument to this view is that insulinomas may be wholly embedded in the pancreas and not visible on its exposed surface. Moreover, they may be difficult to distinguish by palpation from a normal lobule of pancreas or a peripancreatic lymph node and many centres do not have experience in or access to intraoperative ultrasound. For these reasons, an attempt at preoperative localization is still considered good practice, if not mandatory once the definite diagnosis of hyperinsulinism has been made.

Angiography remains a reliable method of delineating insulinomas. The use of superselective injection of contrast with subtraction technique and magnification allows confident identification of tumour 'blush' due to the hypervascularity. Care must be exercised to avoid false-positive localization related to the presence of accessory spleens, large peripancreatic lymph nodes and hypervascular segments of normal pancreas. The calcium arteriogram appears to be the most sensitive test. Feeding pancreatic arteries are selectively catheterized. Calcium stimulates insulin release. A catheter positioned in the hepatic vein is used to measure insulin levels 30–60 seconds following calcium injections. This permits a regional localization.

Recent improvements in imaging technology have enabled increasing detection rates as high as 70% for CT, 80% for MRI and 90% for EUS. Rapid sequence angio-CT allows improved definition of small slightly hypervascular tumours relative to normal pancreas. In general, however, lesions less than about 7–8 mm are poorly detected by currently available imaging tests.

Octreoscan is a non-invasive test in which intravenously injected octreotide, labelled with radioactive tracer, binds to tumours with somatostatin receptors. The sensitivity is related to expression of a special subset of somatostatin receptors. In contrast to other neuroendocrine tumours, insulinomas do not express significant amounts of the subtype 2 (sstr2) somatostatin receptor and for this reason it does not have an established role in the preoperative localization of insulinomas in view of its low yield.

EUS is the latest modality used to detect pancreatic tumours and can identify lesions as small as 2–3 mm. The sensitivity for EUS ranges from 80% to >90% for the detection of insulinomas, but it is a highly observer-dependent technique that demands special expertise.

In spite of their great utility in the localization of these small tumours, the results of imaging tests cannot be interpreted as definitive information. When a solitary adenoma is defined, the surgeon must remember that a second tumour not seen may still be present. Likewise, when several tumours are defined, the probability of additional lesions not seen is very real. Hence, positive preoperative localization studies in no way remove the surgeon's obligation to carry out a complete exploration of the entire pancreas preferably aided by intraoperative contact ultrasound scanning of the entire gland.

When imaging techniques fail to identify the lesion, measurement of immunoreactive insulin in blood sampled from selective catheterization of small pancreatic veins via the percutaneous transhepatic route is indicated. This technique is known as THPVS and is reported to correctly localize tumours not detected by other tests. At the time of portal sampling, samples are also drawn from the hepatic veins to detect metastatic or rare primary sources from within the liver. In addition, arterial samples are drawn periodically to detect any potentially confusing variations in systemic concentrations. THPVS is not without risks and should be confined to centres with very experienced angiographers and to patients who are candidates for abdominal re-exploration following an unsuccessful prior operation.

Surgical management

Without positive preoperative localization of a suspected insulinoma, the surgeon must be totally convinced that the diagnosis is correct before embarking on an operative search. However, with sufficiently strong biochemical evidence supporting the presence of an insulinoma, exploratory surgery is always indicated unless the patient cannot withstand the procedure. The entire pancreas and peripancreatic area must be examined visually and by palpation at operation. This requires access, exposure, assistance and gentle technique. Full mobilization of the gland should always be performed so that careful palpation between thumb and fingers is possible. Intraoperative contact ultrasonography has proved to be of great value in finding small, deeply situated tumours (Figure 27.26).

Solitary insulinomas should be enucleated whenever possible as a good cleavage plane is usually easily established between tumour and adjacent normal pancreas. Care must be taken to avoid injury to the pancreatic duct. Since the great majority of insulinomas are solitary and benign, distal pancreatectomy or, very occasionally, a Whipple type pancreaticoduodenectomy with pyloric preservation is justified for deeply situated tumours that cannot be safely enucleated. Very rarely is a Whipple operation justified for multiple tumours in the head of the pancreas since the likelihood of occult additional tumours being present in the body and tail of the gland is very substantial. Subtotal distal resection for multiple tumours throughout the gland, as seen in MEN-1 patients, is appropriate.

When malignant disease is encountered which can be extirpated by total pancreatectomy and regional lymphadenectomy, this should be done. Even if the tumour is inoperable, as much tumour mass is removed as is safely possible since debulking may provide good palliation with resolution of hypoglycaemic symptoms and increased efficiency of chemotherapy.

With a negative exploration, management options depend upon the clinical situation and the informed consent obtained preoperatively. If not contraindicated by these considerations, it is appropriate to perform pancreatectomy distal to the superior mesenteric vessels. If immediate examination of the thinly sliced resected specimen reveals no tumour and the patient's blood glucose exhibits no rise within 1 hour, it may be elected to

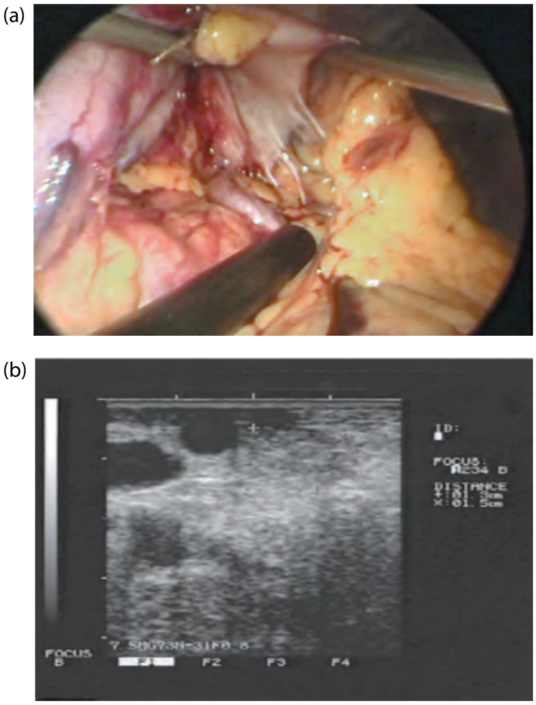


Figure 27.26 Laparoscopic contact ultrasonography of the pancreas after complete exposure of the gland in a patient with an 'occult' 6 mm insulinoma (seen between crosses) and lying close to the splenic vein. To palpation, insulinomas characteristically are slightly firmer than normal pancreas. Enlarged lymph nodes in the peripancreatic region and any liver lesions found should be submitted for frozen section histological evaluation to exclude metastatic disease. Histological examination of primary endocrine lesions in the pancreas is unreliable in the detection of malignancy unless obvious perineural or vascular invasion is present.

perform 90% or even total pancreatectomy. Arguments in favour of the lesser procedure include:

- a small tumour may be overlooked and subsequently found in the resected specimen
- symptoms of hypoglycaemia can be controlled by this procedure alone (in the absence of tumour resection) in 20% of cases
- diazoxide therapy is often successful in controlling symptoms in those uncontrolled by the operation
- the procedure usually does not cause permanent diabetes mellitus.

The benefits of blind total pancreatectomy include:

- removal of an occult lesion that could be an early malignancy
- elimination of the possible need for re-exploration that is difficult and hazardous
- omission of drug therapy for prolonged periods with its undesirable side effects.

It should be mentioned that patients not evaluated preoperatively with percutaneous THPVS for insulin levels perhaps should be referred to a specialized centre for this test and then a second, more directed operation performed. Blind Whipple operations are likewise not recommended.

The results of surgical treatment of insulinoma suggest that about 75% of patients are cured with some 10% developing diabetes following extensive pancreatic resection. About 10–15% of patients have persistent or recurrent hypoglycaemia requiring reoperation at some time. The overall surgical mortality is between 0% and 10% and is related to the extent of resection and expertise of the surgeon. The major operative complications are pancreatitis, abscess, fistula and pseudocyst formation.

Neonatal and infantile hyperinsulinism

Excessive insulin secretion accounts for 20–30% of all cases of unremitting hypoglycaemia in neonates and infants. Such hypoglycaemia can lead to irreversible central nervous system damage and thus requires early recognition, thorough investigation and expeditious treatment. A high intravenous and/or oral glucose intake is mandatory. Additional treatment with diazoxide, somatostatin or a variety of other agents [epinephrine (adrenaline), diphenylhydantoin, glucocorticoids, glucagon, growth hormone] is often required. When hypoglycaemia due to documented hyperinsulinaemia cannot be adequately controlled with medical therapy, urgent operation must be undertaken. Since the overwhelming majority of neonates and infants with hyperinsulinism have nesidioblastosis (-cell adenomatosis or islet cell hyperplasia) as the cause, imaging techniques such as arteriography, ultrasonography and CT have no place in the evaluation of these cases. Likewise, palpation of the pancreas at operation and biopsies for frozen section histological examination are non-contributory. The appropriate procedure is 80–90% extended distal pancreatectomy with splenic preservation. If careful postoperative monitoring of glucose levels indicates inadequacy of the resection, medical therapy should be reinstituted and consideration given to reoperation in cases of further unremitting hypoglycaemia. Reoperation consists of near total (95%) pancreatectomy with preservation of the distal bile duct and duodenum. Permanent exocrine or endocrine insufficiency is unusual in infants less than 3 months of age. This relates to the considerable regenerative capacity of the infantile pancreas.

Medical treatment

Antisecretory therapy with diazoxide is indicated for persistent hypoglycaemia in the preoperative phase and when operation is unsuccessful or contraindicated because of the poor condition of the patient. Diazoxide is a non-diuretic benzothiadiazine which inhibits the release of secretory granules from normal islet -cells and from insulinoma cells. Dosage is individualized based on effectiveness. Because of side effects such as oedema, bone marrow depression, hyperuricaemia, cardiomyopathy and hirsutism in females, patients on diazoxide require close medical supervision. Long-acting analogues of somatostatin are used in the treatment of hyperinsulinism in inoperable patients with insulinoma. Somatostatin both inhibits secretion of peptide by the hyperfunctioning islet cells or tumour and reduces target organ receptivity. In insulinoma patients, two-thirds obtain good symptomatic relief despite much lower rates of control of insulin levels. Few or no antitumour effects have been demonstrated with the use of somatostatin in malignant insulinoma. Streptozotocin, an antibiotic which selectively destroys pancreatic islet cells by inhibiting DNA synthesis, is the chemotherapeutic agent of choice for metastatic insulinoma. Objective tumour regression occurs in about 60% of patients and survival is doubled in those who respond to the drug. Streptozotocin is highly nephrotoxic and hepatotoxic and is thus not used as routine adjunctive therapy. Combinations of somatostatin, streptozotocin and diazoxide are often useful in treating functioning malignant insulinoma.

Gastrinoma (Zollinger–Ellison syndrome)

In 1955, Zollinger and Ellison described two patients, each having a syndrome consisting of fulminant intractable peptic ulcer disease, massive gastric acid hypersecretion, and a non-β-islet cell tumour of the pancreas. Although the same clinical triad had been reported previously, Zollinger and Ellison postulated that the gastric acid excess was caused by a humoral factor released from the tumour. While their original supposition had been that glucagon was the responsible factor, the peptide hormone gastrin was subsequently extracted from such tumours which were found to be of the non-β, non-α-cell type. A radioimmunoassay was developed for gastrin and the hormone was found to be markedly elevated in the plasma of patients with Zollinger–Ellison syndrome.

Clinical features and diagnosis of gastrinoma

The incidence of Zollinger–Ellison syndrome by best estimates is approximately one in 100 000, although the exact incidence is impossible to determine since no large population has been screened for the disease. The disease is more common in men than women with the male-to-female ratio being 3:2. The Zollinger–Ellison syndrome has been reported in patients ranging in age from 7 to 90 years, but the majority of patients have been diagnosed between the third and fifth decades. Only approximately one in 750 patients with the peptic ulcer disease will have gastrinoma as the aetiology.

About one-quarter of Zollinger–Ellison syndrome patients have their gastrinoma as part of the MEN-1 syndrome. MEN-1 is inherited as an autosomal dominant syndrome and the lesions most commonly associated with gastrinoma are parathyroid hyperplasia and pituitary prolactinoma. Gastrinomas in MEN-1 patients are less likely to be malignant but are almost always multifocal. This is in contrast to patients with sporadic gastrinoma in whom the disease is more often malignant but somewhat less frequently multifocal in origin. Overall, the gastrinoma is malignant in about one-half of patients and arises in the pancreas in about 75%. Even when benign, gastrinomas are more often multiple than solitary. The most common extrapancreatic primary tumour site is the duodenum. Tumours in this location are solitary in about one-half of cases. Much less commonly, primary gastrinomas are found in the omenta, lymph nodes, liver and gastric antrum. Malignant gastrinomas metastasize to regional lymph nodes and liver.

Peptic ulcer disease is present in over 90% of gastrinoma patients. Almost all patients with ulcers have typical dyspeptic pain which is more severe and less responsive to medical treatment than in routine peptic ulcer disease. Coexisting diarrhoea is a significant complaint in about one-third of gastrinoma patients. About 5–7% of patients have diarrhoea as their sole presenting complaint. Large volumes of watery stools may result in dehydration, potassium loss, weakness and wasting. The diarrhoea is of the secretory variety and frequently accompanied by steatorrhoea. A multifactorial aetiology has been elucidated but the basic underlying abnormality is acid hypersecretion. With the accompanying rapid gastric emptying, the large acid load in the duodenum and upper jejunum lowers the pH to cause inactivation of pancreatic lipase and other enzymes. This, in addition to the mucosal injury imparted by the large acid load, leads to malabsorption and steatorrhoea. There is also increased intestinal motility and inhibition of salt and water absorption from the jejunum due to the hypergastrinaemia. When severe, the mucosal injury of the distal duodenum and proximal jejunum manifests as frank peptic ulceration at these atypical sites.

The majority of patients with Zollinger–Ellison syndrome are diagnosed only after several years of symptoms, although increasing awareness of the disease is lowering the time between symptom onset and diagnosis. In the past, the majority of patients were diagnosed only after one or more failed operations for presumed routine peptic ulcer disease. Currently, many patients are being diagnosed prior to being subjected to ill-fated standard ulcer operations. All of the complications of peptic ulcer disease are encountered in Zollinger–Ellison syndrome patients and acute haemorrhage and perforation are each noted in approximately 20%. Vomiting and other symptoms of gastric outlet obstruction are distinctly less common as the fulminant nature of the ulcer diathesis more often precipitates acute complications. Wilfred Sircus of Edinburgh referred to patients with Zollinger–Ellison syndrome as 'recurrent ulcerators, persistent perforators, and bleeders unto death.' Severe, refractory reflux oesophagitis has been an underappreciated manifestation of the disease.

Patients with gastrinoma most often have no abnormal physical finding. However, signs of weight loss, epigastric tenderness and intestinal hypermotility are relatively common. Intra-abdominal tumour masses are rarely palpable but hepatic enlargement secondary to massive metastatic deposits is occasionally seen at initial presentation.

The diagnosis of Zollinger–Ellison syndrome should be considered in any patient having:

- severe peptic ulcer disease refractory to histamine H_2-receptor antagonists
- multiple peptic ulcers or ulcers in unusual locations such as the distal duodenum or jejunum
- peptic ulcer disease associated with diarrhoea
- recurrent peptic ulcer disease following an acid-reducing operation
- peptic ulcer disease in association with a strong family history of ulcer disease of MEN-1 syndrome
- peptic ulcer disease without prior diagnosis of MEN-1 syndrome but in association with any other component of MEN-1 syndrome (e.g. hypercalcaemia).

In addition, findings of large gastric mucosal folds and diffuse inflammation or frank ulceration distal to the duodenal bulb on endoscopic or radiological examination of the upper gastrointestinal tract are suggestive of the Zollinger–Ellison syndrome. In the presence of peptic ulceration and/or secretory diarrhoea, radioimmunoassay of the fasting plasma gastrin level remains the key to diagnosis of Zollinger–Ellison syndrome. A basal gastrin level greater than 100 pg/mL strongly supports the diagnosis of gastrinoma. The majority of patients have fasting levels greater than 200 pg/mL and not infrequently 1000 pg/mL or greater. It must be remembered that hypergastrinaemia can occur in association with gastric hypochlorhydria or achlorhydria from a variety of conditions not associated with gastrinoma. Patients with pernicious anaemia, chronic atrophic gastritis, gastric cancer, prior vagotomy or histamine H_2-receptor antagonist or omeprazole therapy may manifest hypergastrinaemia as a physiological response to an elevated antral pH. Thus, it is important to measure basal acid output in all patients suspected to have gastrinoma based on clinical presentation and an elevated plasma gastrin level.

The principal circulating form of gastrin in patients with gastrinoma is G-34 or 'big gastrin', a situation analogous to insulinomas and other peptide-producing endocrine tumours where elevated levels of precursor forms of the respective hormones are often found. The measured circulating gastrin level does not reflect the degree of gastric acid stimulation by the tumour, nor is there good correlation between gastrin level and tumour mass. Plasma gastrin levels normally rise following a meal and thus measurements must be made in the fasting state. Gastric outlet obstruction secondary to ordinary duodenal ulcer disease, antral G-cell hyperfunction or hyperplasia, and retained gastric antrum after Billroth II gastrectomy are other conditions associated with peptic ulcer in which elevated basal plasma gastrin levels may be found (Box 27.4).

In order to differentiate these entities from gastrinoma and also to establish the diagnosis of gastrinoma in ulcer patients with borderline elevated basal gastrin levels, a number of provocative tests have been devised. The best of these is the secretin stimulation test. Following intravenous injection of secretin (2 U/kg), the plasma gastrin level rises within 5–10 minutes to a level 200 pg/mL greater than the basal level in patients with gastrinoma, but not in those with other conditions. The calcium stimulation test has also been used to differentiate gastrinoma from ordinary peptic ulcer disease. In this test calcium is infused at 5 mg/kg/h for 3 hours and a positive test requires that the stimulated gastrin increases by 100% over basal level. Because of untoward side effects of hypercalcaemia, long duration of the test and slightly lower accuracy, the calcium infusion test has been almost entirely supplanted by the secretin injection test. The meal provocation test may be used to differentiate gastrinoma from antral G-cell hyperfunction. In this test, a standard meal is ingested by the patient and causes a marked rise in plasma gastrin levels in those with G-cell hyperfunction but no rise or only a minimal one in gastrinoma patients.

BOX 27.4 Differential diagnosis of hypergastrinaemia

With acid hypersecretion

- Zollinger–Ellison syndrome
- Retained gastric antrum after Billroth II gastrectomy
- Antral G-cell hyperplasia
- Gastric outlet obstruction

With acid hyposecretion

- Pernicious anaemia
- Atrophic gastritis
- Gastric cancer

In order to secure the diagnosis of gastrinoma in patients demonstrated to have hypergastrinaemia, gastric acid secretory testing is necessary. A basal level output greater than 15 mmol/h strongly supports the diagnosis, as does a value greater than 5 mmol/h in the patient with previous acid-reducing gastric surgery for peptic ulcer disease. A ratio of basal acid output to maximal acid output following stimulation with pentagastrin, histamine or betazole of greater than 0.6 has also been used as a discriminatory criterion for gastrinoma. However, this ratio is no more sensitive or specific than is the basal acid output alone. Upper gastrointestinal endoscopy and a standard barium upper gastrointestinal radiological series should be performed in all patients thought to harbour a gastrinoma. In addition to the mucosal abnormalities often found with these studies, on rare occasions a duodenal or antral polypoid lesion has proved to be a gastrinoma on biopsy.

All patients diagnosed as having Zollinger–Ellison syndrome should be further investigated for the presence of MEN-1 syndrome. In addition to the plasma gastrin determinations, several serum calcium and phosphate measurements and a plasma prolactin level constitute the absolute minimal work-up for the most commonly associated endocrine lesions. A family history of refractory ulcer disease, hyperparathyroidism or other endocrine lesions warrants endocrinological screening of the immediate family as well as the patient. The diagnostic approach to the patient with suspected gastrinoma is summarized in Figure 27.27.

Tumour localization techniques

Preoperative techniques to localize gastrinomas are not particularly successful. Since gastrinomas may be small and deeply embedded in the pancreas, abdominal ultrasonography, CT and MRI have not been very sensitive. Recent information, however, suggests that 30–50% of gastrinomas can be identified with multidetector helical CT scanners and EUS. Nonetheless tumours smaller than 7 mm are not usually visualized. Since most gastrinomas are hypovascular, visceral angiography has been much less useful in localizing gastrinomas than insulinomas.

Both EUS and octreoscan have been useful in detection of gastrinomas and other neuroendocrine tumours. The combination of both increases the sensitivity. The sensitivity of EUS is endoscopist dependent but reported detection rates range from 70% to 94%, significantly higher than CT or MRI scanning. Octrescan has its limitations; a high density of somatostatin sstr2 receptors is a necessity. Other tumours or tissues may create a false-positive result. Active infection, inflammation or recent abdominal surgery limits the utility of octrescan. However, these diagnostic tests, if available, can be invaluable preoperatively.

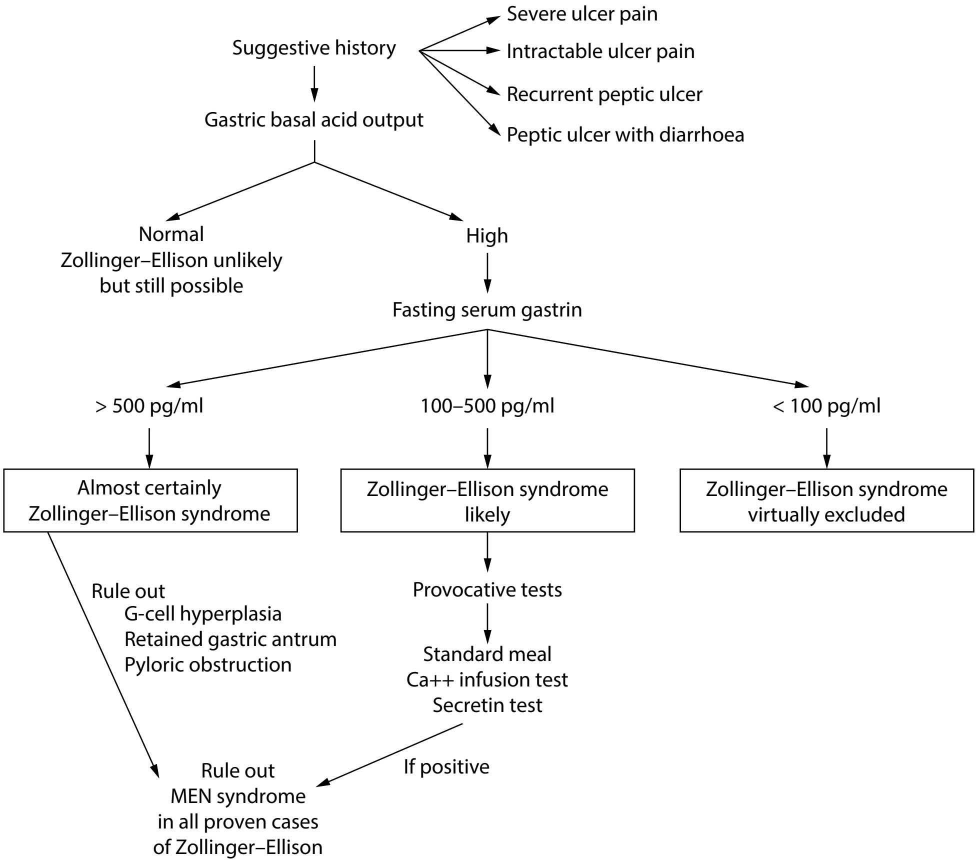


Figure 27.27 An approach to the investigation of patients with suspected Zollinger–Ellison syndrome.

The best results on the preoperative localization of gastrinomas (pancreatic and duodenal) reported in recent years have been with EUS. In one series of 22 patients with Zollinger–Ellison syndrome who underwent exploratory laparotomy after preoperative attempts at localization of the gastrinoma(s) by CT, upper gastrointestinal endoscopy and EUS, the last was by far the most effective. In this study, surgical exploration included intraoperative ultrasonography and duodenal transillumination. The sensitivity of EUS was 50% for duodenal wall tumours, 75% for pancreatic tumours and 62.5% for metastatic lymph nodes. The combination of conventional endoscopy and EUS provided correct preoperative diagnosis in 60% of the patients. This and other reports indicate that EUS should be considered as the first choice of imaging for preoperative detection of gastrinomas.

Percutaneous transhepatic selective sampling of the portal venous system for gastrin levels has been touted to be of value in gastrinoma patients. In theory, the technique is intended to localize the source or sources of hypergastrinaemia and, when combined with hepatic venous gastrin level sampling, predict the location of tumours and the presence or absence of hepatic metastases. Since up to 85% of gastrinomas have been found to reside within the anatomic gastrinoma triangle containing the duodenal C-loop, the head of the pancreas and their associated regional lymph nodes, transhepatic portal venous sampling can only be expected to locate tumours within the triangle but not to localize them more precisely. Therefore, the test is of limited practical utility for gastrinoma.

In general, presently available preoperative tests do not appear capable of localizing the tumour any more reliably than careful intraoperative exploration by the experienced surgeon. Intraoperative contact ultrasonography and transillumination through a flexible endoscope (for duodenal tumours) are valuable in the localization of small tumours. Together with intraoperative contact ultrasound scanning, this should be a routine part of the exploration and may obviate the need for duodenotomy in some cases.

Surgical treatment

In view of their rarity, gastrinomas should be treated in specialized tertiary referral centres. Their management depends on whether they are associated with MEN-1, some of which are treated medically, or are sporadic gastrinomas (not associated with MEN-1) which should always be considered for surgery. In the latter instance, a thorough preoperative assessment of the location and extent of disease is essential (Figure 27.28).

With the advent of effective acid suppression especially by proton pump inhibitors, the role of surgery in the management of patients with Zollinger–Ellison syndrome has changed from surgical treatment of the severe ulcer diathesis resulting from persistent inappropriate acid hypersecretion to improvement of long-term survival by addressing the predominantly malignant nature of primary gastrinomas (90%) despite their slow growth. Total gastrectomy is seldom performed except in exceptional circumstances: patients unable to attend

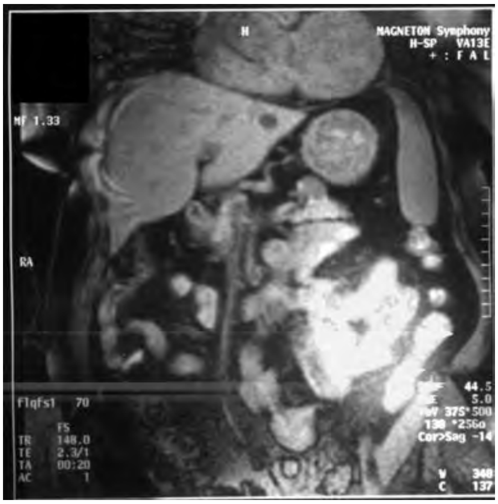


Figure 27.28 Preoperative MRI showing a large pancreatic (2.0 cm) gastrinoma with solitary deposit in the liver. The patient remained free of disease and without need of acid suppression therapy after pancreaticoduodenectomy and hepatic segmental resection.

routine medical follow-up or those with proven failed medical acid suppression or non-compliance with oral medication.

In patients with Zollinger–Ellison syndrome and MEN-1 with hypercalcaemia due to hyperparathyroidism, parathyroidectomy is indicated before any other surgical procedure is contemplated. Parathyroidectomy reduces the serum gastric level in these patients and improves efficacy of medical acid suppression therapy. Surgical resection of gastrinoma is indicated in all patients with sporadic Zollinger–Ellison syndrome who are fit for surgery. Indeed there is some evidence that surgical resection may alter the natural history of the disease since only a small percentage of patients with resected lesions develop liver metastases during follow-up, as distinct to those treated medically. Acid suppression with proton pump inhibitor therapy is established before surgery. At laparotomy a thorough exploration of the pancreas and duodenum (complete Kocherization) is done together with contact ultrasonography and duodenal transillumination via intraoperative upper gastrointestinal endoscopy in the search for the gastrinoma(s). Most surgeons would also perform a duodenotomy if a pancreatic lesion is not found. The search should also extend to examination of regional lymph nodes and metastasis in the liver. Gastrinoma in the pancreatic head is treated by enucleation when possible but a pancreaticoduodenectomy is indicated if the lesion is large or when encucleation would compromise the integrity of the main duct of Wirsung. Hepatic deposits should be excised by appropriate anatomic resection (usually segmentectomy) at the time of resection of the gastrinoma. The long-term outcome of Zollinger–Ellison syndrome patients following curative surgical treatment (Ro) for sporadic gastrinoma (pancreatic or duodenal) is good with 34% being free of disease at 10 years and an overall 10 year survival greater than 80%. A persistently normal postoperative fasting serum gastrin significantly predicts cure. The most important predictor of disease-free survival is histological lymph node status. Reoperation is indicated in patients with sporadic Zollinger–Ellison syndrome who develop persistent or recurrent disease. The site of recurrent disease has to be identified at reoperation (pancreatic, duodenal, metastatic nodal disease of liver deposits) and the disease resected or debulked whenever possible.

The role of surgical treatment in patients with Zollinger–Ellison syndrome associated with MEN-1 is less standardized. Cure is unlikely in these patients because of the high incidence of multiple tumours (80%) and metastatic deposits in lymph nodes at the time of diagnosis of Zollinger–Ellison syndrome. Thus relapse after surgery occurs in the vast majority of patients within 3–5 years. Nonetheless, the prognosis is still good in these patients with or without surgical treatment. The current recommendation is to operate on patients with MEN-1 and Zollinger–Ellison syndrome with large tumour (2.5 or more) as these patients are likely to have metastatic disease.

In view of the above, there is consensus on the following management for newly diagnosed gastrinoma patients.

Patients with preoperatively documented liver metastases and/or MEN-1 syndrome are treated medically in the first instance, especially patients with multiple liver deposits or patients with MEN-1 who have multiple primary tumours. Total gastrectomy is performed if these patients fail medical therapy. During this operation confirmation of the multifocal nature of the disease is obtained by biopsies of suspected tumours.

Young to middle-aged patients without known liver metastases are subjected to the tumour localization and offered elective laparotomy with the primary intent of complete radical excision including regional lymph nodes. This exploration requires full mobilization of the pancreas and the duodenum. All identified lymph nodes and palpable masses within and around the pancreas are removed and submitted for frozen section histology. Blind distal pancreatectomy is not advocated as most tumours occur in and around the head of the pancreas.

Patients responding poorly to antisecretory drugs, or who are non-compliant, and at exploration no tumour is found or multifocal disease is present are treated by elective total gastrectomy which should carry a low mortality (under 5%).

Medical treatment

Since the majority of gastrinomas are malignant, surgical therapy directed at the primary and metastatic disease is indicated in patients with sporadic Zollinger–Ellison syndrome, even if acid suppression is effective.

Currently proton pump inhibitors (omeprazole, lanzoprazole, etc.) have largely replaced H_2-antagonists (cimetidine, ranitidine, famotidine) in acid suppression therapy for the Zollinger–Ellison syndrome. These patients require higher doses (two to four times) that of patients with peptic ulcer disease or gastro-oesophageal reflux. The dose has to be established for the individual patient and not infrequently (25% of patients) increase in dose is necessary over time. Medical acid suppression alone is used more often in patients with Zollinger–Ellison syndrome associated with MEN-1 syndrome with multiple tumours and poor general health.

Treatment of hepatic metastases

Hepatic metastatic disease is usually managed with surgical resection, but other modalities have been used, e.g. hepatic regional arterial infusion and radiofrequency ablation. Liver transplantation may be considered in patients with extensive liver disease but without evidence of extrahepatic deposits in view of the slow growth of these tumours. Somatostatin receptor-directed therapies such as octreotide and radiolabelled octreotide have been used but the reported results are variable.

Systemic chemotherapy is usually with cisplatin and Adriamycin or streptozotocin and 5FU as the last resort in patients with inoperable disease but the results are poor.

VIPoma (Werner–Morrison syndrome; WDHA syndrome; pancreatic cholera)

The syndrome of watery diarrhoea, hypokalaemia and achlorhydria (WDHA) in association with an islet cell tumour of the pancreas was initially described by Werner and Morrison in 1958. A number of hormones have been identified in these tumours (including secretin, glucagon, gastric inhibitory polypeptide, PP and gastrin). However, VIP has been convincingly shown to be the causative agent in the majority of cases. There are occasional patients, however, who have the identical clinical syndrome but with normal plasma VIP levels. In these cases, prostaglandin E is the most likely aetiological candidate. Because this condition resembles cholera, it is also known as pancreatic cholera. A small percentage of patients also have hypercalcaemia, hyperglycaemia, hypochlorhydria and flushing. The condition may be associated with MEN-1. VIP stimulates pancreatic, intestinal and gallbladder water and electrolyte secretions as well as pancreatic enzyme secretion and colonic potassium secretion. VIP inhibits absorption of water and electrolytes (sodium, chloride, bicarbonate and potassium) in the small intestine and colon and also inhibits acid and pepsin secretion in the stomach.

These tumours are usually intrapancreatic, although ectopic primary sites (liver and jejunum) can occur in about 10% of patients. In children, the VIPoma syndrome may be caused by a ganglioneuroma or ganglioneuroblastoma. The diarrhoea in patients afflicted with WDHA is secretory in nature, profuse and causes severe dehydration and loss of potassium. Acidosis almost invariably attends the hypokalaemia and patients suffer from lethargy and weakness as a result of the dehydration and electrolyte abnormalities.

Pathology

VIPomas are composed of sheets of small round cells with uniform nuclei and cytoplasm and rare mitotic figures. They contain electron-dense granules which are due to various tumour products: amines, neuron-specific enolase, synaptophysins, α- and β-subunits of human chorionic gonadotrophin, and chromogranins A, B and C. These tumours frequently are multihormonal, i.e. containing more than one hormone. Unlike insulinomas, VIPomas are generally large in size (>3.0 cm) at the time of clinical presentation, with 50% being malignant though slow growing. The primary tumour is located in the body and tail of the pancreas in approximately 75% of cases and is almost always solitary. The extrapancreatic VIPomas include ganglioneuromas and neuroblastomas and are capable of causing the identical clinical syndrome. Unlike pancreatic tumours which also secrete PP, the extrapancreatic tumours do not.

The pathology of WDHA syndrome is related to the actions of VIP which is normally secreted from non-β-islet pancreatic cells in response to food containing fat, proteins and alcohol. VIP relaxes smooth muscles of the gastrointestinal tract (decrease in lower oesophageal sphincter pressure, relaxation of the gastric antrum and body, inhibition of gallbladder and intestinal circular muscle contraction). VIP has many pharmacological actions: positive inotropic action on the heart; vasodilatation; increase in intestinal water and electrolyte secretion; inhibition of gastrin and gastric acid secretion; and stimulation of pancreatic secretion, lipolysis and glycolysis. Studies have implicated several genes, including the *MEN1* gene, *p16/MTS1* tumour suppressor gene, *DPC4/Smad 4* gene; amplification of the *HER2/neu* proto-oncogene and a possible tumour suppressor gene on chromosome arm 3p. Alterations in the *MEN1* gene and the *p16/MTS1* tumour suppressor gene seem to be particularly involved in the pathogenesis of VIPoma. There is evidence that *p16/MTS1* tumour suppressor gene alterations located on chromosome arm 9p21 occur in a significant percentage of pancreatic endocrine tumours including VIPomas. These mutations result in loss of cell cycle inhibition.

Clinical features

VIPomas are rare pancreatic islet cell tumours, with an estimated incidence of 0.05–0.2 per million population. Their incidence is one-eighth that of insulinomas. The age at diagnosis has a bimodal distribution: children (from 10 months to 9 years) and adults (32–81 years). Diarrhoea is the most common symptom and occurs in at least 90% of patients. The diarrhoea can be associated with abdominal cramps and is voluminous with stool output >3 L/day. It is secretory in nature and persists during fasting. Weight loss occurs in the majority of patients. Abdominal pain is present in 50% and flushing is observed in 20% of patients. Apart from signs of chronic ill health and volume depletion, there are usually no physical findings except for flushing episodes. The definitive diagnosis of WDHA syndrome requires:

- the presence of secretory diarrhoea
- stool volume greater than 700 mL per day
- identification of a pancreatic endocrine tumour
- fasting VIP levels above 190 pg/mL.

Hypokalaemia is present in 90–100% of patients and is secondary to the heavy faecal losses of potassium. Patients often have coexisting hypomagnesaemia and non-anion gap metabolic acidosis, reflecting the severity of the diarrhoea. Hyperglycaemia occurs in 25–50% and is secondary to the glycogenolytic effect of VIP on the liver. Hypercalcaemia occurs in 25–76%. About 6% of these patients have associated MEN-1 syndrome with resultant hypercalcaemia secondary to hyperparathyroidism. Tumours causing hypercalcaemia are almost always malignant. All patients have stool volumes greater than 700 mL/day, and often much higher. The diagnosis of WDHA is suspect if the stool volume is less than 700 mL/day. The diarrhoea in WDHA syndrome is secretory with a stool osmotic gap of less than 50 mEq/L.

Tumour localization

Conventional imaging studies detect fewer than 60% of primary tumours. SRS using indium-diethylene triamine pentacetic acid-phenylalanine (In-DTPA-DPhe) octreotide is the most sensitive modality for identifying the primary tumours and metastatic disease. However, some 12% of positive SRS scans are falsely positive due to the presence of somatostatin receptors in normal tissues. Additional imaging with MRI and selective angiography is however helpful in defining the location of liver metastases. PET is useful for tumour localization, and, when used with ^{18}F-fludeoxyglucose, it may predict malignancy in poorly differentiated tumours. EUS is extremely useful and is considered complementary to SRS scanning in the investigation of patients. If the SRS scan is negative, EUS is performed as it can detect intrapancreatic tumours greater than 0.5 cm. Selective angiography with hepatic venous sampling is indicated when both SRS and EUS are negative. Intraoperative contact ultrasound scanning is recommended during surgical exploration because it may help identify additional tumours that were not detected preoperatively.

Medical treatment

In the first instance, medical treatment is directed at treating symptoms caused by the persistent hypersecretion of VIP with vigorous rehydration and correction of electrolyte and acid–base abnormalities. This can be life saving, since the most common cause of death in VIPoma patients is acute renal failure associated with hypokalaemia. Somatostatin is used to control diarrhoea and may reduce the tumour size in a minority of patients. Somatostatin offers rapid symptomatic control in 90% of patients with marked amelioration of the diarrhoea, dehydration and their attendant metabolic sequelae. Typically, VIP levels decrease but do not fall to within the normal range, suggesting that the attenuation of symptoms is due in considerable degree to reduced target receptivity to the hormone. Slow-release preparations, e.g. lanreotide and sandostatin LAR; are used in preference to octreotide nowadays. Antidiarrhoeal medication such as loperamide may be used in addition to somatostatin in patients with breakthrough diarrhoea. Once the situation of the patient is stabilized, investigation to identify and locate the tumour is carried out if the serum VIP level is elevated.

Surgical treatment

Surgical resection offers the only chance for cure, but the tumour has often spread to regional lymph nodes and/or to the liver by the time of diagnosis. Even so palliative resection of the primary tumour with removal of regional lymph nodes and resection of hepatic metastases can provide substantial palliation. Surgical resection is the only potential cure as 50–60% of VIPomas are malignant. Thus all fit patients should be considered for exploratory laparotomy. The operation starts as an exploration during which the liver is examined for metastatic disease. Next a careful examination of the entire pancreas is carried out with complete Kockerization of the duodenum and incision of the retroperitoneum along the superior and inferior aspects of the pancreas. Mobilization of the spleen may be required in order to examine the pancreatic tail. Additionally, the small bowel should be explored for extrapancreatic tumours. Although VIPomas are usually intrapancreatic, 10% of tumours are ectopic along the ganglia of the autonomic nervous system.

Intraoperative contact ultrasound is extremely useful and may identify lesions not detected preoperatively. Isolated tumours in the pancreatic body or tail are removed by distal pancreatectomy, with or without splenectomy. More proximal tumours in the pancreatic head or body may be treated by enucleation. However, the most common operation performed for proximal tumours is pancreaticoduodenectomy.

Patients with resectable disease have a 5 year survival rate as high as 80%, as distinct from patients with incompletely resected/ or unresectable disease in whom the 5 year survival drops to 25%. However, cytoreductive/debulking surgery may prolong life expectancy. Liver transplantation for metastatic liver disease is considered in patients who are fit and free of extrahepatic disease.

Palliative chemotherapy

Patients who are beyond treatment by surgical resection, or who recur with multiple metastatic deposits, may have symptomatic responses to long-acting somatostatin analogues. Streptozotocin is reasonably effective in palliation but, immediately following administration, diarrhoea and electrolyte losses may be exacerbated for several days, necessitating aggressive replacement. The prognosis is poor with advanced disease, and thromboembolic complications due to excessive dehydration add to the morbidity and mortality. The occasional patient with pancreatic cholera secondary to prostaglandin E2 hypersecretion may have dramatic relief with indometacin therapy.

Aside from streptozotocin, several chemotherapeutic agents have been used to treat patients with metastatic disease but streptozotocin and doxorubicin are the most effective. Other agents include 5FU and a combination of interferon-α and 5FU. Radiologically guided therapies include hepatic artery occlusion, with or without chemotherapy, chemoembolization with doxorubicin in iodized acid combined with gelatin or sponge particles.

Glucagonoma

Glucagonomas are tumours of the A-cell of the pancreatic islet and are responsible for a characteristic syndrome consisting of severe skin rash, weight loss, diabetes mellitus, deep venous thrombosis, anaemia and hypoaminoacidaemia. The tumour is very rare and because its most salient feature, the skin rash, is frequently misdiagnosed, its true significance and the correct diagnosis are often made very late in the course. As expected, the disease is virtually always first seen in the dermatology clinic. Glucagonoma is considerably more common in females and is a disease of middle age. At the time of diagnosis approximately 60–70% of cases have already metastasized. The most common site of metastasis is the liver (50%). The typical clinical features of the glucagonoma syndrome are shown in Table 27.7.

Table 27.7 Clinical features of glucagonoma syndrome

Sex	Female/male >1
Age range (mean)	20–71 (57)
Malignancy	70–80%
Clinical diabetes (abnormal glucose tolerance test)	70% (95%)
Skin lesions	80%
Glossitis/cheilitis	90–100%
Weight loss	90–100%
Diarrhoea	50–60%
Coarse intestinal mucosal folds	50–60%
Anaemia, normocytic, normochromic	70–80%
Hypoaminoacidaemia	80%
Neurological deficit	Unusual (incidence uncertain)
Marked hyperglucagonaemia	100%
Survival	80% more than 2 years
	23% more than 5½ years

The typical skin rash is termed necrolytic migratory erythema. The lesions are characteristically symmetrical and erythematous with crusted erosions involving the perineum, groins, thighs, buttocks and distal extremities. The systemic manifestations of weight loss, weakness and lethargy are due to a combination of the catabolic effects of high plasma glucagon levels and the extensive malignant disease which is often present. Hyperglycaemia results from increased hepatic glycogenosis and gluconeogenesis. Most patients are frankly diabetic but ketonaemia rarely develops because circulating insulin levels are increased. Panhypoaminoacidaemia is a uniform finding and may be responsible for the skin rash as well as the neurological deficits that are occasionally seen. The anaemia is characteristically normocytic and normochromic and, although serum iron levels may be low, the anaemia does not respond to iron and vitamin replacement.

The diagnosis is usually made upon recognition of the typical cutaneous manifestations in combination with diabetes mellitus in the setting of a chronic wasting disorder. The diagnosis is confirmed by finding an elevated plasma glucagon level. Normal values range between 50 and 150 pg/mL. Values of glucagon greater than 1000 pg/mL are often seen with glucagonoma but interpretation should be cautious in the absence of the typical skin rash or other suggestive signs. The numerous other conditions associated with hyperglucagonaemia include:

- diabetes, ketoacidosis, hyperosmolar syndrome
- chronic renal failure
- shock states:
  - myocardial infarction
  - septicaemia
  - burns
  - hypovolaemia/haemorrhage
- acute pancreatitis
- cirrhosis
  - portosystemic shunting (natural, surgical)
- glucagonoma
- familial hyperglucagonaemia (asymptomatic)
- exercise
- antiglucagon antibodies in diabetics treated with insulin.*

In most glucagonoma patients glucagon release from the tumour can be induced by the administration of arginine or tolbutamide. PP levels are elevated in one-half of patients. Since many cases are diagnosed in an advanced metastatic stage, tumour localization is not normally difficult. CT, angiography, EUS, MRI and octreotide scintigraphy have all been used successfully in localizing tumours. While topical steroids and intravenous amino acid administration have been effective in ameliorating the skin eruption in some patients, definitive treatment is surgical. Operative exploration is indicated even for advanced metastatic disease as debulking procedures may significantly alleviate the debilitating catabolic effects of the excess glucagon. When surgical resection is not an option, selective arterial embolization and chemotherapy are indicated. Streptozotocin combined with 5FU can produce a reduction in both tumour size and glucagon levels. Dimethyltrizenoimidazole carboxamide has been effective in providing symptomatic relief and alleviation of skin rash. Somatostatin has proved highly efficacious both in the preoperative management and as palliative therapy in glucagonoma. Octreotide or, more usually, its long-acting analogues reduce circulating glucagon levels, dramatically improve the skin rash, attenuate the systemic symptoms and augment the anabolic effects of intravenous hyperalimentation. Use of the somatostatin analogues is indicated in all cases regardless of the stage of disease or surgical plan. Unfortunately, no effects on tumour size or growth have been demonstrated. Similarly, interferon-α has been used with some success. This induces an autoimmunity against the tumour. This treatment is not curative, but it can prolong survival and control the symptoms of the disease. Combination treatment with interferon-α and somatostatin or chemotherapy has shown synergistic or additive beneficial effects in some patients with glucagonoma and other neuroendocrine malignancies.

Somatostatinoma

Somatostatinoma is a very rare tumour with less than 90 cases reported to date. The patients have been mostly middle-aged. The patients with pancreatic somatostatinoma are predominantly female. Over 80% of the tumours have been associated with liver metastases at the time of diagnosis. Along with somatostatin production most tumours elaborate other hormones in addition, such as VIP, PP, gastrin, calcitonin or cortisol.

Despite the potent inhibitory nature of somatostatin, the usual clinical syndrome is non-specific. Abdominal pain is the most common presenting symptom and this may relate to the high prevalence of cholelithiasis in these patients. Gallbladder stasis is thought to be of aetiological importance. Other symptoms and signs commonly associated with somatostatinoma are diarrhoea, diabetes mellitus (25%), weight loss, anorexia, hypochlorhydria, steatorrhoea and anaemia. Symptoms not related to excessive somatostatin levels such as tachycardia, flushing, hypertension,

---

* These may be induced by slight glucagon impurities in commercially available insulins and manifest as hyperglucagonaemia when glucagon is measured by a double antibody immunoassay method.

hypokalaemia and hypoglycaemia are also present in some patients. The diagnosis is made by chance in most cases although a radioimmunoassay for plasma somatostatin is available. Somatostatin released by the tumour may be stimulated by tolbutamide; however, the utility of this provocative test is unknown at present.

Ideally, treatment is by surgical excision of the pancreatic or duodenal lesion. Debulking of advanced tumours may be efficacious and some patients have benefited from adjunctive therapy with streptozotocin and 5FU. The 5 year survival rate after diagnosis of somatostatinoma is only about 15%.

Pancreatic polypeptide tumour (PPoma)

PP was discovered in 1972 as a single major protein from a crude insulin preparation. In mammals, 93% of the cells producing PP are located in the pancreas and are known as F or D2 cells. There is an association between increased PP levels and increased intra-abdominal but not subcutaneous fat, as measured by CT scan. A clear-cut biological role for PP has not been established. The only physiological effects recognized in humans are the inhibition of gallbladder contraction and pancreatic enzyme secretion. Thus, a tumour originating from PP cells is likely to be clinically silent. High levels of PP reduce both appetite and food intake in normal healthy volunteers indicating that PP may be a satiety hormone. PP is also secreted by other islet cell tumours in addition to their specific hormones, especially after stimulation by a meal. PP is thus now considered as a useful adjunct marker in the detection of islet cell tumours. PP-producing tumours are mostly located in the pancreas and may present as three pathological lesions: (1) pure PPomas, (2) mixed tumours with minor PP cell population and (3) PP-cell hyperplasia. These tumours are among the most common multiple adenomas found in patients with multiple endocrine neoplasia type 1. Numerous types of extrapancreatic endocrine tumours are able to synthesize and secrete PP. They occur mostly but not exclusively in the gastrointestinal tract, particularly in the rectum. The inactivation of the *MEN1* gene at 11q13 appears to be involved in the development of pancreatic but not of rectal PP-producing tumours.

Basal concentrations of PP in plasma may be raised above 1000 pg/mL in 22–77% by all endocrine-secreting tumours but are invariably normal in patients with ductal adenocarcinomas of the pancreas. The accuracy of raised basal PP concentration as a marker for endocrine-secreting tumours can be increased by determining the response of PP to secretin administration: a response of greater than 5000 pg/min/mL is considered abnormal (greater than two standard deviations observed in healthy subjects). Many cases of so-called non-functional endocrine tumours are now thought to be PPomas, because 50–75% of them have raised basal PP levels and an exaggerated response to secretin. Thus, in the absence of chronic renal failure which causes marked elevation of PP, markedly raised PP levels in otherwise healthy patients may indicate a non-functioning pancreatic endocrine tumour. Atropine suppression of PP has been suggested as a method of determining whether the elevated hormone is the result of tumour secretion.

PPomas are rare and are not associated with a clinical syndrome, despite the fact that they are functional and secrete high levels of PP. The tumours are usually single and benign, although a few have metastasized to the liver. There are now at least 25 patients in the literature with PPomas with an age spanning 20–74 years (mean of 51 years) and with an equal sex incidence. Diarrhoea occurs in one-third of cases, whereas steatorrhoea is universal in patients in whom it is sought. Decreased acid secretion is rare but 60% of patients present with weight loss. Some patients exhibit a pruritic rash. Otherwise the PPoma syndrome is silent, and these tumours often discovered during investigation of patients with hepatomegaly (metastases), abdominal pain, obstructive jaundice and haematochezia. Upper gastrointestinal bleeding may occur due to invasion of the wall of the duodenum or thrombosis of the splenic or portal vein, with consequent development of varices. PPomas are recognized by arteriography as highly vascular tumours with or without metastases to the liver. Twenty-five per cent of PPomas are associated with MEN-1 syndrome. In general however PPomas have a relatively benign clinical course with a low recurrence rate following excision at 5 years. The frequency of malignancy of PPomas is not established.

It has been suggested that every patient with a markedly elevated level of PP should undergo exploratory laparotomy with careful inspection of the pancreas. If a tumour is identified it is excised and the resection and extent will depend on site and size of the lesion. This practice is disputed by others as, although rare, PPomas may occur in the chest and elsewhere outside the pancreas. With this treatment policy which appears more logical, somatostatin receptor scintigraphy and EUS should be performed preoperatively to locate the tumour in patients with elevated PP. The findings of these imaging tests then dictate whether a laparotomy or other exploration is necessary. Metastatic PPomas are best treated with streptozotocin plus doxorubicin.

Multiple endocrine neoplasia type 1 syndrome (MEN-1; MEA-1; Werner syndrome)

The MEN-1 syndrome is inherited as an autosomal dominant disorder but considerable phenotypic variability exists even within an individual family. However, the pancreas, parathyroid glands and pituitary are involved in all patients if examined pathologically. The pancreas is inevitably involved with diffuse islet cell disease consisting of micronodular and macronodular hyperplasia, and often multiple tumours secreting multiple peptide hormones.

The parathyroid glands are most frequently involved in MEN-1 syndrome with hyperparathyroidism being present in 85% of cases. In the vast majority of these all four glands are affected. This is in contrast with the very low incidence of parathyroid hyperplasia in isolated primary hyperparathyroidism. Pancreatic abnormalities occur in over 80% of MEN-1 patients, with non- -cell tumours being most common. The most common pancreatic tumour found in MEN-1 syndrome patients is gastrinoma, and in virtually all such patients multiple pancreatic tumours are found.

Of the pituitary lesions, chromophobe adenomas and particularly prolactinomas are the most frequent lesions encountered. When small, these tumours may be without symptoms but, in male patients, they are associated with manifestations of the antiandrogenic effect of prolactin. Tumours producing growth hormone and leading to acromegaly are the next most frequently encountered variety.

While the MEN-1 syndrome is classically associated with lesions of the parathyroid glands, pancreatic islets and pituitary gland, an increasing number of adrenocortical lesions have been recognized. Many of these lesions are non-functioning adenomas; however, glucocorticoid excess has been reported in some. Other occasional associations of MEN-1 include thyroid nodules, bronchial and intestinal carcinoids and lipomas.

All patients with endocrine pancreatic tumours should be carefully investigated for additional manifestation of MEN-1 syndrome. Thus, estimations of serum levels of calcium and phosphate as well as plasma assays for parathormone, insulin, gastrin, glucagon, somatostatin, PP, prolactin, growth hormone, ACTH and cortisol constitute a relatively complete though by no means exhaustive endocrine evaluation for the MEN-1 syndrome. Whenever a patient is diagnosed as having MEN-1 syndrome, screening of all available family members is indicated.

Multiple endocrine neoplasia type 2 syndrome (MEN-2; MEA 2; Sipple syndrome)

MEN-2 is another discrete endocrine syndrome which is inherited as an autosomal dominant syndrome with variable expressivity. It is not associated with pancreatic disease. It consists of hyperparathyroidism, medullary carcinoma of the thyroid gland and phaeochromocytoma. MEN-2b is a variant which is also inherited as an autosomal dominant syndrome but, unlike in MEN-2, the incidence of parathyroid disease is extremely low. The syndrome consists of multiple mucosal neuromas, intestinal ganglioneuromatosis leading to megacolon and constipation, a Marfanoid habitus and characteristic facies with thickened lips and alae nasi, along with medullary carcinoma of the thyroid and phaeochromocytoma.

GUIDE TO FURTHER READING

Albazaz R, Da Costa PE, Verbeke CS. Pancreatic polypeptide cell hyperplasia of the pancreas. *J Clin Pathol* 2006;**59**:1087–1090.

Aljarabah M, Ammori BJ. Laparoscopic and endoscopic approaches for drainage of pancreatic pseudocysts: a systematic review of published series. *Surg Endosc* 2007;**11**:1936–1944.

Asakawa A, Inui A, Yuzuriha H, *et al.* Characterization of the effects of pancreatic polypeptide in the regulation of energy balance. *Gastroenterology* 2003;**124**:1325.

Batterham RL, Le Roux CW, Cohen MA, *et al.* Pancreatic polypeptide reduces appetite and food intake in humans. *J Clin Endocrinol Metab* 2003;**88**:3989–3992.

Beger HG, Krautzberger W, Bittner R, *et al.* Duodenum preserving resection of the head of the pancreas in patients with severe chronic pancreatitis. *Surgery* 1985;**97**:467–473.

Bouvet M, Gamagami RA, Gilpin EA, *et al.* Factors influencing survival after resection for periampullary neoplasms. *Am J Surg* 2000;**180**:13–17.

Bradley EL III, Reynhout JA, Peer GI. Thoracoscopic splanchnicectomy for 'small duct' chronic pancreatitis: case selection by differential epidural analgesia. *J Gastrointest Surg* 1998;**2**:88–94.

Büchler MW, Gloor B, Müller CA, *et al.* Acute necrotizing pancreatitis: treatment strategy according to the status of infection. *Ann Surg* 2000;**232**:619–626.

Burns GP, Bank S (eds). *Disorders of the Pancreas.* New York, NY: McGraw-Hill, 1992.

Cuschieri A, Shimi SM, Crosthwaite, G, *et al.* Bilateral endoscopic splanchnicectomy through a posterior thoracoscopic approach. *J R Coll Surg Edinb* 1994;**39**:44–47.

De-Madaria E, Soler G, Martinez J, *et al.* Update of the Atlanta classification of severity of acute pancreatitis: should a moderate category be included? *Pancreatology* 2009;**9**:433–434.

Dowen SE, Cronogorac-Jurcevic T, Gangeswaren R, *et al.* Expression of S100P and its novel binding partner S100PBPR in early pancreatic cancer. *Am J Pathol* 2005;**166**:81–92.

Etemad B, Whitcomb DC. Chronic pancreatitis: diagnosis, classification, and new genetic developments. *Gastroenterology* 2001;**120**:682–707.

Frey CF, Smith GJ. Description and rationale of a new operation for chronic pancreatitis. *Pancreas* 1987;**2**:701–707.

Isenmann R, Runzi M, Kron M, *et al*, for the German Antibiotics in Severe Acute Pancreatitis Study Group. Prophylactic antibiotic treatment in patients with predicted severe acute pancreatitis: a placebo-controlled, double-blind trial. *Gastroenterology* 2004;**126**:997–1004.

Kuo SC, Gananadha S, Scarlett CJ, *et al.* Sporadic pancreatic polypeptide secreting tumors (PPomas) of the pancreas. *World J Surg* 2008;**32**:1815–1822.

Lall CG, Howard TJ, Skandarajah A, *et al.* New concepts in staging and treatment of locally advanced pancreatic head cancer. *AJR Am J Roentgenol* 2007;**189**:1044–1050.

Leung TK, Lee CM, Lin SY, *et al.* Balthazar computed tomography severity index is superior to Ranson criteria and APACHE II scoring system in predicting acute pancreatitis outcome. *World J Gastroenterol* 2005;**11**:6049–6052.

Lytras D, Manes K, Triantopoulou C, *et al.* Persistent early organ failure: defining the high-risk group of patients with severe acute pancreatitis? *Pancreas* 2008;**36**:249–254.

Moossa AR, Robson MC, Schmipff SC (eds). *Comprehensive Textbook of Oncology*, 2nd edn. Baltimore, MD: Williams & Wilkins.

Mortenson M, Bold RJ. Symptomatic pancreatic polypeptide-secreting tumor of the distal pancreas (PPoma). *Int J Gastrointest Cancer* 2002;**32**:153–156.

Mozell E, Stenzel P, Woltering EA, *et al.* *Functional Endocrine Tumors of the Pancreas: Clinical Presentation, Diagnosis, and Treatment. Current Problems in Surgery.* Chicago, IL: Year Book, 1990.

Neoptolemos JP, Kemppainen EA, Mayer JM, *et al*. Early prediction of severity in acute pancreatitis by urinary trypsinogen activation peptide: a multicentre study. *Lancet* 2000;**355**:1955–1960.

O'Toole D, Salazar R, Falconi M, *et al*. Rare functioning pancreatic endocrine tumors. *Neuroendocrinology* 2006;**84**:189–195.

Petrov MS. Meta-analyses on the prophylactic use of antibiotics in acute pancreatitis: many are called but few are chosen. *Am J Gastroenterol* 2008;**103**:1837–1838.

Petrov MS, Windsor JA. Classification of the severity of acute pancreatitis: how many categories make sense. *Am J Gastroenterol* 2010;**105**:74–76

Petrov MS, van Santvoort HC, Besselink MG, *et al*. Enteral nutrition and the risk of mortality and infectious complications in patients with severe acute pancreatitis: a meta-analysis of randomized trials. *Arch Surg* 2008;**143**:1111–1117.

Puestow CB, Gillesby WJ. Retrograde surgical drainage of the pancreas for chronic pancreatitis. *Ann Surg* 1958;**76**:898–906.

Ravi K, Britton BJ. Surgical approach to insulinomas: are pre-operative localisation tests necessary? *Ann R Coll Surg Engl* 2007;**89**:212–217.

Sharma VK, Howden CW. Prophylactic antibiotic administration reduces sepsis and mortality in acute necrotizing pancreatitis: a meta-analysis. *Pancreas* 2001;**22**:28–31.

Tong J, Utzschneider KM, Carr DB, Zraika S. Plasma pancreatic polypeptide levels are associated with differences in body fat distribution in human subjects. *Diabetologia* 2007;**50**:439–442.

Tonsok K, Takamichi M, Satoru T, *et al*. Pancreatic CT imaging: Effects of different injection rates and dose of contrast material. *Radiology* 1999;**212**:219–225.

Vege SS, Gardner TB, Chari ST, *et al*. Low mortality and high morbidity in severe acute pancreatitis without organ failure: a case for revising the Atlanta classification to include 'moderately severe acute pancreatitis'. *Am J Gastroenterol* 2009;**104**:710–715.

Working Party of the British Society of Gastroenterology; Association of Surgeons of Great Britain and Ireland; Pancreatic Society of Great Britain and Ireland; Association of Upper GI Surgeons of Great Britain and Ireland. UK guidelines for the management of acute pancreatitis. *Gut* 2005; **54**(Suppl. 3):1–9.

CHAPTER 28

Bariatric surgery

AHMED R. AHMED AND CAREL L.G. ROUX

Obesity

The problem of obesity: obesity-related diseases, life expectancy and health costs

The World Health Organization (WHO) estimated that globally in 2011 approximately 2 billion adults were overweight and at least 500 million were obese. In almost all developed countries and in many developing ones, obesity rates have risen dramatically. According to the Information Centre for the NHS in 2011, 24% of adults (aged 16 years or older) in England were classified as obese. This has nearly doubled since 1993. Women are more likely to be morbidly obese (3% compared with 1% of men). The implications are that 21% of males and 23% of females are classed as being at very high risk of obesity-related health disease.

Obesity and being overweight can have a variety of adverse health consequences associated with a high rate of premature death, such as type 2 diabetes, dyslipidaemia, hypertension, obstructive sleep apnoea, certain types of cancer, steatohepatitis, gastro-oesophageal reflux, arthritis, polycystic ovary syndrome (pcoss) and infertility. The life expectancy of a severely obese person is reduced by an estimated 5–20 years. An estimated 2.6 million people die each year as a result of being overweight or obese and the consequences of excess weight will soon rival tobacco as the world's leading cause of preventable premature deaths.

The obesity pandemic also has significant economic consequences. The WHO estimates that, in many developed countries, obesity now accounts for 2–7% of all healthcare spending. Yet medical costs are only a small fraction of the pandemic's total costs. Among its other adverse economic effects are heightened absenteeism rates, reduced worker productivity, and increased food and clothing costs.

Pathophysiology of obesity

The laws of thermodynamics plainly state: energy will be stored if more energy is consumed than expended. The storage is usually in the form of adipose tissue and it has been calculated that the average man living in the USA increases his weight by more than 9.1 kg between the ages of 25 and 35. This dramatic weight accumulation is due to only a 0.3% imbalance between energy consumed and expended during this period. A suggestion that the obesity pandemic is due to a more sedentary lifestyle is based on the fact that several studies could not find a significant increase in the amount of calories consumed over the last few decades. The suggestion that the current situation can be remedied by increasing exercise levels seems to be prone to failure for several reasons. First, the human body is extremely efficient when it comes to energy expenditure. The average man who finishes a marathon in 4.5 hours will burn only 2072 kcal, approximately the same amount of energy he would normally be expected to consume during a day. Even allowing for the increased energy expenditure that follows exercise, he would have to run a marathon every 10 days while not increasing his energy intake if he hoped to lose 1 kg a month. Second, our daily surroundings have been optimized to allow us to spend the minimum amount of energy. This includes a reduced energy expenditure required, for example using the motor car instead of public transport, while we also use escalators and lifts instead of stairs when we reach our destinations.

Appetite is a complex process controlled by several neural, humoral and psychological factors. It can be argued that in the past there has never been an evolutionary advantage in weight loss and as such we have evolved to consume as much food as possible whenever we can. A powerful drive to eat is induced by the homeostatic system regulating energy balance, especially after weight loss. Other basic drives include breathing and quenching one's thirst. Although it is possible for an individual to consciously hold his or her breath the compulsion to take another breath quickly overcomes the best effort. Hunger is an equally potent drive, not as intense as breathing, but probably not dissimilar to the drive to drink when one is thirsty.

The appetite regulation systems have always been under intense selective pressures and a gene–environment interaction is thus likely. Hunter-gatherers had to face famines and could only rely on food being sporadically available. The thrifty gene hypothesis states that genes that increased energy stores and predisposed to obesity could have provided a survival advantage.

Non-surgical treatments: lifestyle, pharmacotherapy and behavioural changes

In general, results of intentional weight loss and weight maintenance following lifestyle changes have been disappointing. A multitude of studies have examined the benefits of a low-calorie diet, with or without exercise and/or behaviour therapy. Most of the studies reported a degree of weight loss, but these were predominantly short-term effects. However, studies reporting longer term follow-up data demonstrated that the excess weight invariably returned to similar or even higher levels. This is consistent with the theory that early humans who existed as hunter-gatherers experienced continuous cycles of feast and famine. This may have resulted in an evolutionary advantage to those that could in times of famine recognize and act upon the signal of hunger. A further advantage could possibly be gained by those also able during times of feast to replace all the weight loss and even increase weight above previous levels to allow some buffer for future times of famine. The inherited ability to regain weight may contribute to the "yo-yo dieting" pattern so commonly observed today.

The role of pharmacotherapy in achieving weight loss is becoming less controversial. Many clinicians still feel that obesity is a function of poor will power and patients should on these grounds not be considered for pharmacotherapy. However, the same argument was previously levelled at hypercholesterolaemia, and it was only after effective agents became available to treat this co-risk factor for cardiovascular disease that pharmacotherapy became more widely accepted. Pharmaceutical companies are investing heavily in drugs to target obesity. What has remained unclear is what part of this complex system that controls energy homeostasis should be targeted to achieve the best results. The remarkable redundancy within the central nervous system that has evolved to ensure prevention or correction of weight loss during starvation with the least detrimental influence on health has prevented many pharmacological agents making it to market.

Currently, only one agent is available. Orlistat, a lipase inhibitor, prevents up to 30% of the absorption of fat by the gastrointestinal system. The mechanism of weight loss is however not directly linked to the pharmacodynamics, but more to the side-effect profile when fatty foods are ingested while orlistat is being used. The resulting diarrhoea or oily leak is unpleasant enough to discourage the continuation of a high-fat diet. The patient, if well informed, will usually change to a lower fat and thus less energy-dense diet, with a resulting decrease in weight. Weight loss on orlistat therapy is not a universal phenomenon, because many perpetual dieters consume low-fat foods. However, the volume and calorific value of these meals still ensure an energy surplus. These patients tend not to do well on orlistat, because unpleasant side effects occur less often due to the reduced fat in their diet.

Drugs such as sibutramine and rimonabant were withdrawn secondary to side-effect profiles that included increased cardiovascular death in patients with some pre-existing risks and significant depression, respectively. Several drugs are currently in development, but the weight loss rarely exceeds 10%. Recently the FDA approved two new drugs Qsymia™ and Belviq™ which are currently undergoing clinical trials. Despite the initial setbacks pharmacotherapy is predicted to continue to show rapid growth as most lifestyle interventions have been shown to be ineffective if not undertaken in conjunction with pharmacotherapy.

Weight loss utilizing surgical interventions

Surgical interventions for weight loss – also referred to as bariatric surgery – have proven to be the most effective method for intentional weight loss and post weight loss maintenance. The rationale for the original interventions was predominantly based on either restricting the progression of food through the gastrointestinal system or causing some form of malabsorption or a combination of the two. Weight loss surgery is not a cosmetic procedure and does not involve the removal of adipose tissue. It has been shown that animals undergoing such procedures have a marked reduction in appetite, a result which has also been observed in humans. The two possible mechanisms causing the reduction in appetite include neural and humoral factors. Most likely, both factors have roles and this hypothesis for weight loss is becoming increasingly popular.

Before patients can be considered for surgery, traditional methods such as diet, exercise and pharmacological agents should have been tried. The indications for bariatric surgery, based on the US National Institutes of Health criteria, are shown in Box 28.1.

These eligibility criteria may need to be re-evaluated as there is strong evidence of the effectiveness of bariatric surgery in ameliorating obese comorbidities independent of body mass index (BMI). The International Diabetes Federation argued that in some high-risk populations, e.g. Southeast Asian, bariatric surgery should be offered at a lower BMI in order to treat diseases like type 2 diabetes.

Undoubtedly, for the correct group of patients, surgery leads to significantly greater, and sustained, weight loss than that achieved with non-surgical treatments. This leads to the resolution of, or significant improvement in, many weight-related diseases and conditions, which in turn leads to increasing a patient's life expectancy and significantly improving their

BOX 28.1 Eligibility criteria for bariatric surgery

- Body mass index (BMI) ≥ 35 kg/m^2 with obesity related comorbidities (such as diabetes, sleep apnoea or joint problems)
- BMI ≥40 kg/m^{2*}
- Have tried all other methods of losing weight (diets, exercise, medication) including 6 months of a medically supervised programme but not been able to sustain weight loss
- Have no specific medical or psychological reasons why they should not have this type of surgery
- Are fit enough to undergo an anaesthetic and surgery
- Understand the need for long-term follow-up

* In the UK, The National Institute for Health and Clinical Excellence now recommends all those with a BMI ≥50 to be considered directly for surgery without the need for a prior 6 month medically supervised weight loss.

quality of life. There have been many high-quality studies in the literature supporting the above: the Swedish Obese Subjects (SOS) study; Sjöström *et al.*, Christou *et al.*, Adams *et al.* and Buchwald *et al.*

As a consequence of the effectiveness of weight loss surgery, there have been large increases in this field of surgery worldwide. In the USA, since 2006, the gastric bypass operation has been more prevalent than laparoscopic cholecystectomy. In the UK, recent data from the National Bariatric Surgery Registry (NBSR) and Hospital Episode Statistics (HES) data also document a 10-fold increase in bariatric operations from 238 operations in 2000 to 2543 in 2007. Despite these increasing trends, only a very small proportion of obese patients are receiving surgery, mainly due to artificial constraints placed by purchasers and non-coverage by insurers. A widening gap exists between those who would benefit from surgery and the lucky few who receive it.

Brief history of weight loss surgery

Jaw wiring was first used in the mid-1970s and had the purpose of restricting the opening of the patient's mouth, thus making the consumption of a usual meal impossible. These patients could therefore only consume liquid meals by using a straw. This highly regimented protocol ensured dramatic weight loss as the liquid diet contained much reduced calories. This form of surgical intervention was accompanied by a behaviour modification programme that tried to teach the patients to follow a low-calorie diet after the removal of the jaw wires. The end results were however very disappointing, although all the patients lost significant amounts of weight while their jaw movement was restricted they invariably regained all the weight after the wires were removed.

The first account of abdominal surgery for massive obesity is by Henrikson in the late 1950s where a massive small bowel resection was performed, leaving a short bowel. This led others to perform bypasses of the small intestine. The jejunoileal bypass (JIB) induced a state of malabsorption by bypassing most of the intestines while keeping the stomach intact. Although the weight loss with the JIB was good, too many patients developed serious complications such as blind loop syndrome due to bacterial overgrowth, abdominal bloating, migratory arthralgias, urinary stones and, unless adequate protein was consumed, liver problems. There were also electrolyte imbalances and anal burning from diarrhoea. Consequently, many patients have required reversal of the procedure.

In order to perform an operation without the side effects of the JIB, in the mid-1960s Mason developed the gastric bypass. This was based on the weight loss observed among patients undergoing partial stomach removal for ulcers. Initially the operation was performed as a loop bypass with a much larger stomach. Because of bile reflux that occurred with the loop configuration, Griffen modified the operation and performed it as a Roux-en-Y with a limb of intestine connected to a stomach pouch which prevents the bile from entering the upper part of the stomach and oesophagus. Torres made a further modification to this – a vertical pouch on the lesser curvature of the stomach. The Roux-en-Y gastric bypass (RYGB) has undergone many modifications since. In 1994, Drs Wittgrove and Clark reported the first case series of laparoscopic RYGB (LRYGB). The laparoscopic technique applies itself effectively to RYGB and its popularity continues to increase primarily because it has fewer peri- and postoperative complications (in particular wound problems), a shorter hospital stay, and a more rapid recovery time than open surgery.

Around the same time as the bypass was being developed, Scopinaro in Italy developed the biliopancreatic diversion (BPD), which was designed to be a safer malabsorptive alternative to the JIB. This operation induces controlled malabsorption without many of the potential side effects caused by bacterial overgrowth associated with the JIB. This primarily malabsorptive operation only allows for absorption of carbohydrates and fats in the distal 50–70 cm of ileum. BPD has been reported to provide the highest long-term weight loss of any bariatric operation. However, patients need close long-term postoperative follow-up for protein, vitamin and mineral deficiency as these remain significant complications following surgery.

The duodenal switch (BPD-DS) was a modification of the BPD proposed by Hess to prevent stomal ulcers, increase the amount of gastric restriction, minimize the incidence of dumping syndrome and reduce the severity of protein-calorie malnutrition. Anatomically, the main difference between the DS and the BPD is the shape and size of the stomach – the malabsorptive component is essentially identical to that of the BPD. Instead of cutting the stomach horizontally and removing the lower half (such as with the BPD), the DS cuts the stomach vertically and leaves a tube of stomach (a sleeve gastrectomy) that empties into a very short (2–4 cm) segment of duodenum. The BPD-DS can be performed using the laparoscopic approach and remains a popular option in a few centres around the world; however, most bariatric units are concerned with the side effects of malabsorption (daily diarrhoea), coupled with the possible complications of protein malnutrition and metabolic bone disease and thus the BPD-DS only forms ≤1% of all bariatric operations performed worldwide.

The first part of the DS (laparoscopic sleeve gastrectomy) is now being performed widely as a sole initial operation in poor-risk patients, with the option to add the DS portion or RYGB as a second operation if there is insufficient weight loss or weight regain. A 2009 consensus meeting on the sleeve gastrectomy reports impressive results with this operation as weight loss and health benefits are in line with those seen after LRYGB. The sleeve gastrectomy is gaining in popularity for these very reasons, but some authors have reported a small proportion of patients regaining weight and requiring further surgery.

In search of a safer, simpler technique, Carey and Gomez designed the gastroplasty in the early 1970s as an alternative to the RYGB and the JIB. The gastroplasty was the first purely restrictive operation performed for the treatment of obesity. However, the small horizontal pouch enlarged as did the outlet. Mason in 1982 reported the vertical banded gastroplasty (VBG) in which a vertical pouch is fashioned along the lesser curve of the stomach with the outlet controlled by a non-dilatable band. The VBG is still being performed with various modifications including via the laparoscopic route. However, its popularity has waned due to frequent regain of weight in the long term resulting from either pouch enlargement or fistula development between the two partitioned segments of stomach.

Molina in the early 1980s developed gastric banding as a potential restrictive procedure. A band was placed high around the stomach, producing a tiny pouch. Subsequently, others developed an inflatable gastric band attached by tubing to a subcutaneous reservoir through which saline could be instilled or withdrawn via a needle in order to adjust its tightness around the stomach, resulting in attenuated hunger and hence a reduction in food intake. The laparoscopic adjustable gastric band (LAGB) remains a popular surgical option worldwide, its success being dependent upon regular band follow-up/adjustments.

In addition to the operative interventions described above, a number of non-surgical procedures have also been developed as a treatment for excess weight. In 1985, a procedure was developed that aims to restrict the food consumed by the placement of a space-occupying device within the gastric lumen deployed via a gastroscope – the intragastric balloon. More recently, the EndoBarrier gastrointestinal liner system is an endoscopically delivered device that offers an alternative. Other techniques using endoscopic devices to produce gastric restriction are being developed, as are different ways to access the peritoneal cavity to perform currently accepted operations (natural orifice transendolumenal surgery and single incision laparoscopic surgery).

Preparation for surgery

Unit experience and infrastructure

The laparoscopic approach is technically challenging and there is a definite learning curve seen for bariatric surgeons. Depending on the reported case series, the number of laparoscopic gastric bypasses performed before levelling of the learning curve ranges between 75 and 100. This is again different for each procedure, with proficiency reported after 50 cases for either a sleeve gastrectomy or a gastric band. Once the curve plateaus, there is an observed improvement in patient outcome with respect to complication rate.

It is essential that any unit offering bariatric surgery should have the correct infrastructure in place. This includes wide corridors, doors, correct beds, trolleys, chairs, wheelchairs and floor-mounted toilets in any area where a bariatric patient may be seen – outpatient, inpatient ward, radiology suite, intensive care, emergency room, etc. Furthermore, the operating room must be adequately equipped for the bariatric patient in terms of the correct operating table, optics and laparoscopic instruments.

Thus the costs of establishing an integrated bariatric programme can be quite expensive as bariatric hospital beds at 2010 prices can cost up to US$30 000 and operating tables up to US$60 000. Furthermore, investment in training healthcare professionals is essential, not only in order to have the right bariatric multidisciplinary team (MDT) but also to ensure that other hospital staff ranging from porters, receptionists and nurses show appropriate sensitivity and care towards the bariatric patient.

The role of the bariatric multidisciplinary team and the patient journey

The input from a psychiatrist, a dietician, a bariatric physician, a specialist nurse and an anaesthetist is vital to promote a good outcome for the patient. Dietician input both before and after surgery is highly valuable so that patients can be informed of the necessary changes in eating (both quantity and type of food) that need to be made. Furthermore, assessment by a psychiatrist helps ensure that those patients with pre-existing psychological comorbidities receive therapy to ensure optimal benefit from surgery. Similarly, those patients with drug addictions, active psychoses and personality disorders, which all represent poor compliance with the surgery and its aftercare, will be filtered out by the psychiatrist. Figure 28.1 illustrates the typical bariatric surgery patient journey from referrer to surgery and thereafter.

Good preoperative work-up prior to a general anaesthetic is crucial. The patient will often have many comorbidities, including diabetes, hypertension, mobility difficulties secondary to arthritis and obstructive sleep apnoea. The last one may be undiagnosed, and if it remains undetected the patient can have anaesthetic complications postoperatively, including respiratory failure. Immobile patients are at high risk of deep vein thrombosis (DVT) and may thus require preoperative inferior vena cava (IVC) filter insertion. Thus care and attention in optimizing patients before surgery will reduce and often eliminate the need for intensive unit care in the peri- and postoperative period as well as reduce postoperative complications.

Certain groups of patients are at higher risk for complications because of the high prevalence of comorbidities, at the same time these very patients may also be the ones who benefit most from interventions. The King's assessment criteria can be used for an objective preoperative assessment of patients (Table 28.1).

Before surgery

To optimize intraoperative conditions, the patient should undertake a specifically designed preoperative diet (approximately 900 kcal/day) for at least 2 weeks. This is in order to allow a fatty liver to shrink in size, hence permitting sufficient visualization and access to the left upper quadrant for the operation to take place.

Correct patient positioning is key for any bariatric procedure, not only in order for the surgeon to gain adequate access, but also for patient safety. Sometimes the operations can be prolonged, putting the patient at significant risk of developing intraoperative complications, including DVT, neuropraxia and rhabdomyolysis. An appropriate operating table with the facility of maximal head-up, with securing straps and footboards, is essential. Calf compression devices should be used during surgery to help prevent DVT, and similarly the use of subcutaneous heparin prior to incision is recommended for the same reason.

Choice of procedure

Traditionally patients who are 'volume eaters' have been recommended to undergo gastric banding whereas those with binge eating, snacking and sweet-eating tendencies have been selected for gastric bypass surgery. The sleeve gastrectomy was used as the first stage in a two-stage operation culminating

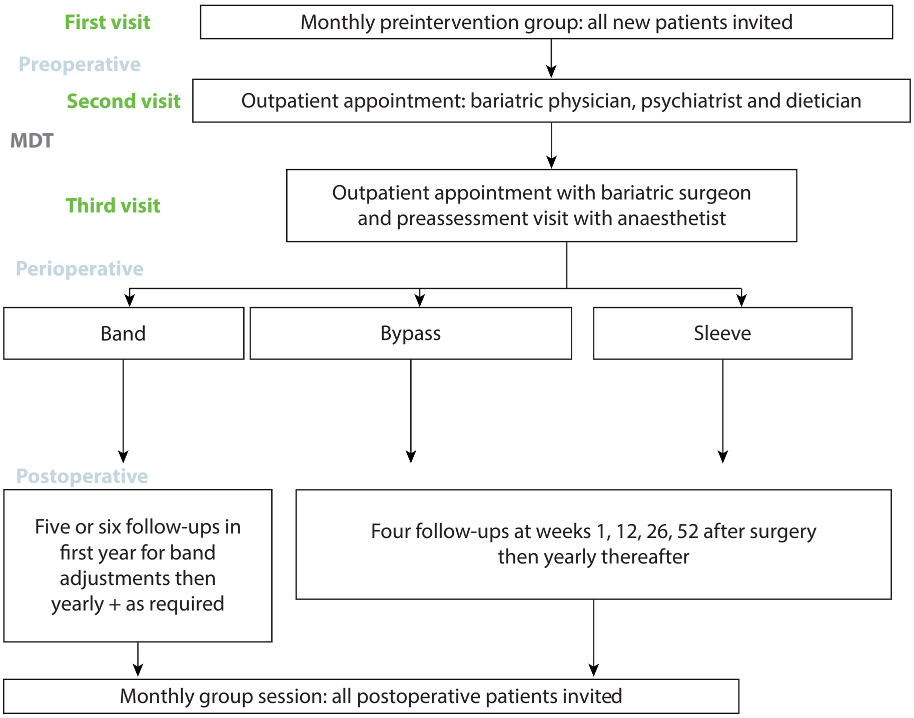

First visit
Monthly preintervention group: all new patients invited
Preoperative
Second visit
Outpatient appointment: bariatric physician, psychiatrist and dietician
MDT
Third visit
Outpatient appointment with bariatric surgeon and preassessment visit with anaesthetist
Perioperative
Band
Bypass
Sleeve
Postoperative
Five or six follow-ups in first year for band adjustments then yearly + as required
Four follow-ups at weeks 1, 12, 26, 52 after surgery then yearly thereafter
Monthly group session: all postoperative patients invited

Figure 28.1 A typical bariatric surgery patient pathway. MDT, multidisciplinary team.

Table 28.1 Modified King's criteria

	Stage 0	Stage 1	Stage 2	Stage 3
Airway	Normal	Snoring	Sleep apnoea require CPAP	Cor pulmonale
BMI (kg/m^2)	<30	30–35	35–50	>50
Cardiovascular	<25% risk	>25% risk	Heart disease	Heart failure
Diabetes	Normal	Impaired fasting glycaemia	Type 2 diabetes	Uncontrolled type 2 diabetes
Economic	Normal	Suffered discrimination	Unemployed due to obesity	Requires financial support
Functional	Can manage 3 flights of stairs	Manage 1 or 2 flights of stairs	Manage less than 1 flight of stairs or requires walking aids	Housebound
Gonadal	Normal	Irregular periods	PCOS/impotence	Infertility
Health status	Normal	Low mood or QoL	Moderate depression or poor QoL	Severe depression
Image	Normal	Does not like looking in mirror	Avoids mirrors/body image dysphoria	Severe eating disorder
Junction gastro-oesophagus	Normal	Heart burn	Oesophagitis	Barrett's oesophagus
Kidney	Normal	Proteinuria	GFR <60 mL/min	GFR <30 mL/min
Liver	Normal	Raised LFT/NAFLD	NASH	Liver failure

BMI, body mass index; CPAP, continuous positive airway pressure; GFR, glomerular filtration rate; LFT, liver function test; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; PCOS, polycystic ovarian syndrome; QoL, quality of life.

with a DS in the super-superobese (BMI >60). However such artificial divisions do not necessarily guarantee the best outcomes and probably what matters the most is that choice of surgical intervention should be made jointly by the patient and the clinician, taking into account the degree of obesity, comorbidities, the best available evidence on effectiveness and long-term effects, the facilities and equipment available, and the experience of the surgeon who will perform the operation. In fact, all bariatric procedures when used in motivated patients in a healthcare system able to provide regular follow-up will achieve their primary objective of improving health.

Patient safety

One of the most significant studies ever conducted on the safety of bariatric surgery is the Longitudinal Assessment of Bariatric Surgery study published in 2009. It followed 4776 patients who had bariatric surgery for the first time, evaluating complications and death rates within the first 30 days after surgery. Thirty-day mortality was low, ranging from no deaths in the laparoscopic adjustable gastric band group to 0.2% in the LRYGB group. Within 30 days of surgery, 4.1% of patients had developed at least one complication. The findings of this research very strongly reaffirm the safety of bariatric surgery.

Common types of weight loss surgery

Laparoscopic gastric band (Figure 28.2)

Surgical technique

A pneumoperitoneum is created in the standard technique and the liver left lobe retracted. Hiatus hernias should be looked for and repaired prior to proceeding with band placement. Traditionally the band was placed via a perigastric approach. The current standard of care is to use a pars flaccida approach, via a retrogastric tunnel between the pas flaccida medially and the angle of His laterally, since this produces a significantly decreased rate of band slippage. The band should be checked and prepped as per the manufacturer's instructions before placing it within the abdomen. The retrogastric tunnel is created by placing a Goldfinger at the right crus fat pad and then by rotating the tip through to the angle of His. If there is a large fat pad then this can be mobilized off the stomach. The band is then attached to the Goldfinger and gently pulled through and locked in position. To help prevent band slippage three or four anterior gastrogastric placation sutures are placed over the anterior aspect of the band. The catheter tubing that leads from the band is connected to a port reservoir. This is implanted in the subcutaneous layer of the abdominal wall over the rectus fascia, which allows adjustment of the band to the desired degree of constriction in the outpatient clinic. The band can be performed in 30–40 minutes as a day-case procedure.

Complications

The gastric band is the safest procedure in terms of operative mortality (<0.1%) but risk of further surgery for a band-related complication is in the range 5–10% over the following 10 years (Table 28.2)

Gastric or oesophageal perforation is extremely rare and is iatrogenic at the time of band implantation. Stoma obstruction can occur at any stage. In the early postoperative phase, it is the result of oedema or haematoma around the band, which often settles with time. Later on it is usually due to poor dietary habit and bolus obstruction. Gastric band erosion signifies slow migration of the band across the gastric wall thus making the band visible within the gastric lumen. The whole band may thus become internalized. The presentation is usually through weight regain and/or port infection. Gastric band slippage results from the stomach inferior to the band herniating above it, thus resulting in excessive stomach tissue inside the band, and to obstruction between the upper pouch and lower stomach. This condition usually presents acutely with vomiting. Oesophageal and gastric pouch dilation refers to symmetrical enlargement of the gastric pouch above the band, which may in turn lead to dilation and malfunction of the oesophagus (megaoesophagus). This entity is associated with maladaptive overeating and aggressive band adjustments. Symptoms include reflux and a decrease in satiety and restriction. However, it is not usually associated with obstruction. Patients may present with a sense of decreased restriction, requesting a band fill. However, this will exacerbate the problem. Pouch dilation can usually be treated non-operatively by temporarily defilling the band but sometimes needs band explantation.

Table 28.2 Gastric band complications

Band-related complication	Incidence	Port-related complication	Incidence
Slippage	4%	Port infection	5–10%
Erosion	1%	Port displacement	
Symmetric pouch or oesophageal dilation	5–10%	Needle-stick injury to tube	
Gastric perforation	0.1%		
Stoma obstruction	Variable		

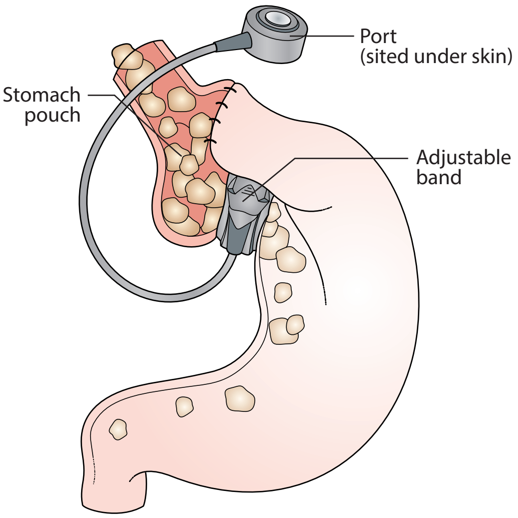


Figure 28.2 Gastric band.

Postulated mechanisms of action

Historically gastric banding was thought of as a restrictive technique; although that may have been the intention, it now appears after work by O'Brien *et al.* that restriction does not predict weight loss. Rather, the reduction in hunger that is a feature in 70–80% of patients with gastric bands predicts weight loss.

The mechanisms that cause the reduced hunger after gastric banding are not yet clear, but the hypotheses favoured at present include pressure on the vagus nerve fibres in the region where the band is situated. The traditional satiety gut hormones such as peptide YY (PYY), glucagon-like peptide 1 (GLP-1) and oxyntomodulin (OXM) do not appear to play a role after gastric banding. The changes in energy expenditure, ghrelin and leptin are also consistent with the weight loss observed after gastric banding. The improvements in glycaemic control depend on the weight loss and occur in a temporally consistent fashion to the changes in peripheral insulin sensitivity.

Results

Weight loss following gastric banding is variable and depends on the unit where it is carried out and patient follow-up. Some units in Australia report weight loss outcomes around 70% of excess weight, matching those seen with gastric bypass surgery. On average, however, most units will achieve around 50–55% excess weight loss (EWL) (roughly equivalent to around 20% total body weight loss) with the band.

Laparoscopic Roux-en-Y gastric bypass (Figure 28.3)

Surgical technique

The gastric bypass operation commences with formation of the vertical gastric pouch some 15–20 mL in volume. Dissection starts along the lesser curve, and the retrogastric adhesions and angle of His are freed. Sequential firings of an endostapler create a vertical gastric pouch. There has been much research into the size of the pouch. A retrospective study has suggested that smaller pouches may be associated with greater weight loss, although accurate measurement of pouch volume is difficult and prospective data are lacking. Most surgeons choose the transection point by measuring from the oesophagogastric junction as accurately as simply "eye-balling" the anatomy, or by counting vascular arcades. If too large, it can cause an increase in the rate of marginal ulceration and decreased weight loss.

In order to create the Roux-en-Y bypass, first the jejunum is divided at a point typically 30–40 cm distal to the ligament of Treitz. The distal segment is then moved cephalad to form the alimentary or Roux limb, and is surgically connected to the gastric pouch. This segment can be brought either ante- or retrocolic, and ante- or retrogastric, and there are different techniques in use for the formation of the gastrojejunostomy (handsewn, linear stapled or circular stapled). Benefits of the antegastric route include the relative increased ease of accessing the anastomosis and re-examining it if necessary at subsequent surgery, particularly in revision or relook surgery since it is not hidden underneath the stomach. Following creation of the gastrojejunostomy, most surgeons check for a leak intraoperatively using either methylene blue dye or oxygen through an orogastric (OG) tube with the anastomosis submerged under saline. A leak is thus ruled out in the absence of bubbling or blue dye. The proximal bowel segment, also called the biliopancreatic limb, is usually connected to the alimentary limb 75–150 cm distal to the gastrojejunostomy.

Several authors have addressed the issue of alimentary limb length during RYGB. For patients with a BMI of 50 kg/m^2 or less, there is no proven benefit for alimentary limbs longer than 150 cm. Other studies examining the use of alimentary limbs longer than 250 cm for patients with a BMI greater than 50 kg/m^2 have found improved weight loss over standard RYGB, but if greater than 300 cm there is increased nutritional deficiencies and need for reoperation.

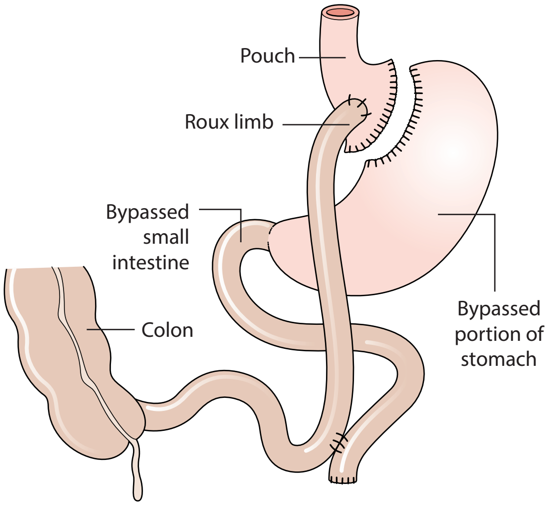


Figure 28.3 Roux-en-Y gastric bypass

One of the consequences of performing a RYGB is the creation of mesenteric defects in the transverse mesocolon (if retrocolic Roux limb) or jejunojejunostomy or through Petersen's defect. The latter defect is between the mesentery of the Roux limb and the transverse mesocolon. Internal hernias are more commonly observed in the laparoscopic RYGB than open, postulated to be a result of the fewer adhesions found in the laparoscopic technique. Closure of these defects at the primary operation is not performed universally by all surgeons, but overall is recommended by most surgeons.

The gastric bypass can typically be performed in 60–90 minutes with a 2 day stay in hospital.

Complications

These can be classified into general complications from bariatric surgery (which may also be seen with all bariatric operations) and more specific ones related to the bypass.

General complications include venous thromboembolic phenomena such as DVT (1%) and pulmonary embolism (0.1%) can be fatal (the International Bariatric Surgery Registry notes this as the most frequent cause of mortality following bariatric surgery) and all efforts are made to prevent these by using subcutaneous heparin peri- and postoperatively, calf compression devices, early mobilization post surgery and, where necessary, a low threshold for placing an IVC filter prior to surgery. Nutritional complications following surgery are unlikely as long as patients are compliant with their intake of multivitamin and mineral supplements. Patients should be followed up yearly and undergo blood tests to screen for micronutrient deficiencies, in particular iron, calcium, vitamin B_{12} and vitamin D deficiency (this may have been present prior to surgery). Other rare findings include vitamins A, B_1 and C, and selenium and copper deficiencies. Subsequent to rapid weight loss, bariatric surgery patients are at a higher risk of developing gallstones and may present with acute cholecystitis, biliary colic, choledocholithiasis and gallstone pancreatitis.

Specific complications related to the laparoscopic gastric bypass include gastrointestinal tract obstruction. This can occur at the gastrojejunostomy from a postoperative stricture (1%) or food bolus obstruction. More distally, small bowel obstruction (SBO) may be related to internal hernia formation (1–2%) where small bowel becomes trapped within iatrogenic gaps in the mesentery of the small bowel or transverse colon (retrocolic LRYGB). A further complication of the retrocolic LRYGB is Roux limb obstruction caused by narrowing within the transverse mesocolic defect. This tends to present earlier than internal herniation and is usually caused by cicatrix formation and extrinsic circumferential compression of the Roux limb. Other possible causes of SBO in this population include intussusception; adhesions; port site hernias; and obstruction at the jejunojejunostomy from kinking, stricture, or blood clot.

Gastric remnant distention can occur acutely or chronically and may present in the early postoperative period or years after surgery from obstruction of the biliopancreatic limb or common channel. Patients are usually in distress and have epigastric pain, nausea and tachycardia. In addition to leucocytosis, patients with an obstruction distal to the second portion of the duodenum may have elevated liver function test results and pancreatic enzymes from high duodenal pressure.

Marginal ulcers may develop at any stage after surgery and may occur in up to 10% of patients. The aetiology is multifactorial and may be related to one or more of the following: gastric acid; tobacco; non-steroidal anti-inflammatory drugs; *Helicobacter pylori*; gastro-gastro fistula; anastomotic tension and/or ischaemia, foreign body (suture) and large pouch size. Patients with ulcer perforation typically experience acute, severe epigastric pain and present with tachycardia, fever, leucocytosis and free air on plain radiographs or CT studies.

In the early phase following surgery bleeding (1–2%) and leaks (1–2%) at any of the staple lines or anastomoses may occur and require urgent reoperation. Delay in the treatment of a leak may result in severe sepsis and death. The presentation of a leak in the bariatric patient may be subtle with few symptoms other than just feeling 'unwell' and a mild tachycardia. Hence the bariatric surgeon must have high vigilance and a low threshold for re-exploration. Generally speaking, acute leaks in sick patients need relaparoscopy/laparotomy, washout and drainage. Late leaks in well patients may be manageable with conservative treatment using antibiotics, percutaneous drainage and stents.

Postulated mechanisms of action

The majority of modern Roux-en-Y gastric bypass techniques do not result in any calorie malabsorption or restriction of food through the gastric pouch. The operation can be viewed as having at least five components which may contribute to the mechanisms resulting in weight loss. These components include the small pouch, the bypass of the remnant stomach and first part of the small bowel, the undiluted bile flow in the first part of the small bowel, the early contact of the mid-jejunum with food and the disruption of the small vagal fibres when the stomach pouch is separated from the remnant stomach.

The mechanisms most prominent after the gastric bypass include the early satiety, reduction in calorie intake, the possible increase in energy expenditure and the changes in food preference from high-fat high-sugar foods to lower fat and low-sugar foods. Aversion or avoidance of sugary foods may seen subsequent to be 'dumping syndrome', which some patients experience after gastric bypass surgery.

The role of the satiety gut hormones PYY, GLP-1 and OXM has received prominence over the last few years, but the increasing focus on the role of bile acids, the neuronal signals generated by the altered anatomy and changes in gut microbiome may yield new insights. The improvements in glycaemic control after gastric bypass appear to be due to the increased insulin secretion as driven by the enhanced incretin responses, the improved insulin sensitivity, which includes both the early changes in hepatic insulin sensitivity and the later changes in peripheral insulin sensitivity as well as the change to a low glycaemic index diet.

Results

Weight loss following LRYGB is rapid in the first 6 months; thereafter, it plateaus and is 70–75% EWL (roughly equivalent to around 30% total body weight loss).

Laparoscopic sleeve gastrectomy (Figure 28.4)

Surgical technique

Following creation of a pneumoperitoneum, the operation starts by entering the lesser sac, then the gastrocolic and gastrosplenic ligaments are divided using an energy device. The lateral attachments of the stomach are divided along its greater curve. The new stomach can be sized around a bougie, or OG tube, the size of which varies between 32 and 44 French gauge. There is some evidence that smaller sized tubes (32–34 Fr) achieve greater weight loss. Formation of the sleeve usually starts 2–10 cm proximal to the prepyloric vein of Mayo. It is currently under debate whether sparing or resecting part of the antrum has an effect on weight loss. The endostapler is fired once via the right port, usually with a larger green or black cartridge, because of the thick-walled gastric antrum. The remaining stapler reloads (purple or gold) are placed from the left port. As the stomach is divided in a sleeve gastrectomy, the greater curve specimen to be removed assumes a diamond shape with a large round central portion tapering to a point at its upper and lower extent. Extraction of the gastrectomy specimen is through one of the laparoscopic ports, which may require gentle widening.

Some surgeons oversew the staple line or use fibrin glue or omentum to help prevent against haemorrhage or leak. Other surgeons use buttressing materials at the time of firing the endostapler. A variety of such materials are available and include: bovine pericardium (Peri-Strips Dry; Synovis Life Technologies, Inc.) and polyglycolic acid copolymer (SeamGuard; W.L.Gore & Associates, Inc.)

The laparoscopic sleeve gastrectomy can be performed in 45 minutes with a 1 day stay.

Complications

Historically the sleeve gastrectomy was used in high-risk patients as the first stage of a two-stage BDP-DS. This high-risk group, characterized by the accepted high-risk factors, including BMI ≥60 kg/m^2, demonstrated a notably low rate of complications including 1% for staple-line leakage and 2% for bleeding

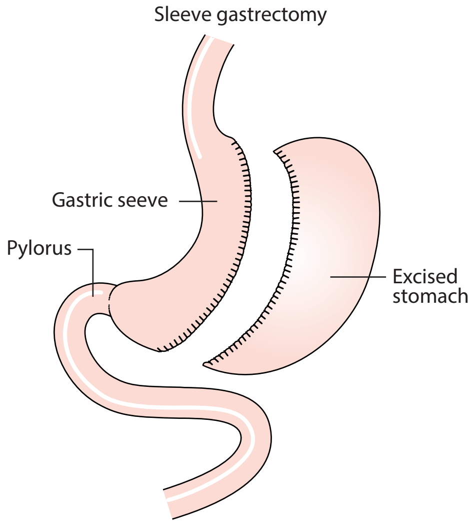


Figure 28.4 Sleeve gastrectomy.

and mortality within 30 days occurring in only 0.2%. With refinement of the sleeve gastrectomy technique as a primary procedure (including the use of staple line reinforcements) it is likely these complication rates will be further reduced.

Similar to other bariatric procedures, long-term nutritional surveillance is recommended especially as long-term postoperative nutritional complications may occur secondary to the extensive gastric resection by decreasing the absorption of some vitamins and nutrients, such as vitamin B_{12} and iron.

Postulated mechanisms of action

The mechanisms by which sleeve gastrectomy work are receiving major interest at present and may help revise many of the dogmas that currently prevail in bariatric surgery. The sleeve was specifically designed to cause restriction, but recent data show no restriction of food through the pylorus. Moreover, rodents with sleeve gastrectomy retain the ability to increase their food intake above levels of sham-operated animals if they are placed in a state where their hunger signals are activated. The satiety gut hormone responses of PYY and GLP-1 are similar to those of a gastric bypass. Changes in food preferences and energy expenditure may be less pronounced than after gastric bypass, but this has not been formally tested. Improved glycaemic control appears to be due to increased insulin secretion, although to a lesser extent than after gastric bypass. Insulin sensitivity also increases, but the time course appears more consistent with peripheral insulin sensitivity improvements related to weight loss than the early hepatic insulin sensitivity changes seen after gastric bypass.

Finally, the role of the pylorus after sleeve gastrectomy requires further study, as a known effect of the satiety gut hormones are to close the pylorus, and it is possible that once calorie intake has reached a certain threshold the pylorus will be closed prematurely resulting in a sudden feeling of fullness. Patients quickly learn not to continue eating once this has occurred as the side effect of vomiting may result.

Results

Weight loss following sleeve gastrectomy has been noted to be in the range of 60–70% EWL (roughly equivalent to around 25% total body weight loss) 1 year after surgery. The main problem is a deficiency of long-term follow-up data in the literature to confirm the effectiveness of this operation as a stand-alone intervention at 5 years or more.

Biliopancreatic diversion and duodenal switch (Figure 28.5)

Surgical technique

Both operations can be performed laparoscopically but take longer due to their technical complexity. Furthermore, interest is waning away from them due to (1) increased risk of serious complications, (2) need for close long-term follow-up to monitor for protein calorie malnutrition and bone mineral disease and (3) loose smelly stools from fat or bile acid malabsorption which may occur 2–4 times/day.

Complications

Table 28.3 lists the complications associated with the BPD and BPD-DS.

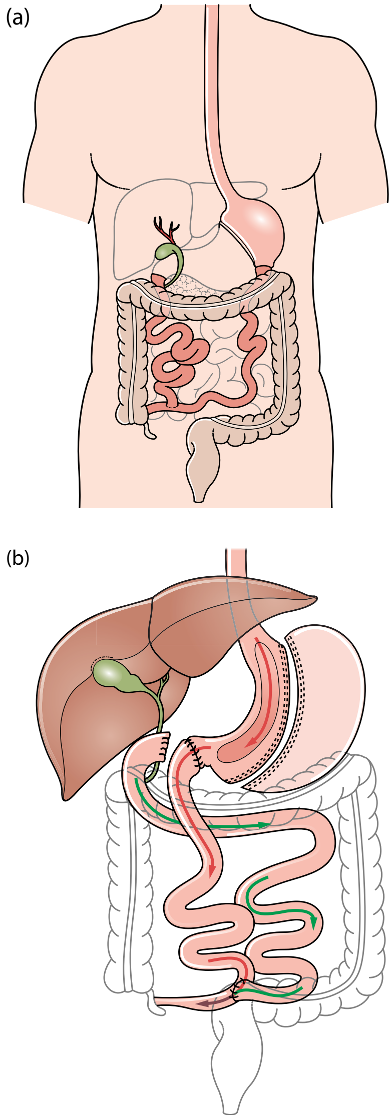


Figure 28.5 (a) Biliopancreatic diversion and (b) duodenal switch.

Postulated mechanisms of action

Calorie malabsorption is a feature of BPD and BPD-DS, but often the calorie loss in the faeces does not adequately explain the reduction in weight. It should however be noted that some patients have a very large increase in food intake after BPD and this may be secondary to the calorie malabsorption where the patient tries to compensate for the calorie loss by increasing the calories consumed.

The usual findings after BPD or BPD-DS are however a reduction in calorie intake and an increase in energy expenditure. The changes in food preference are less well documented. The improvements in glycaemic control appear to be primarily due to the improvements in insulin sensitivity. The BPD is the only procedure studied thus far that reduces peripheral insulin sensitivity early. This may be due to the reduction in intramuscular lipid content. Understanding the role of bile acids in the much increased biliopancreatic limb may help elucidate the mechanisms of the BPD and BPD-DS.

Table 28.3 Complications associated with BPD and BPD-DS

Complication	Incidence (%)
Protein malnutrition	12
Anaemia	35
Incisional hernia (if performed open)	10
Stomal ulcer (less with BPS-DS)	5
Bone demineralization	15
Vitamin deficiency (A, D, E, K, and β-carotene)	30
Dumping syndrome (less with BPD-DS)	

BPD, biliopancreatic diversion; DS duodenal switch.
Source: American Society of Bariatric Physicians.

Results

The BPD and BPD-DS offer the least reduction in eating capacity with the most weight loss with 80–90% EWL (equivalent to 40% total body weight loss) as well as the highest rates of resolution of the metabolic syndrome. However, because of the complex nature of the surgeries, risk of complications and many side effects, these operations remain less popular with most bariatric surgeons.

A summary of the common bariatric operations in current clinical usage is shown in Table 28.4.

Other bariatric interventions

Intragastric balloon

This is implanted under endoscopic guidance and fills the stomach and thus reduces the potential volume for food in the stomach. The presence of the balloon will cause severe vomiting if too much food is consumed. The device is temporary and needs to be removed after 6 months. Although indicated as a primary treatment for those who are overweight (BMI >27) and those suffering from obesity, this procedure may also be used to prepare extremely obese patients for other forms of weight-reducing surgery. The average weight loss seen with the balloon is around 15–20 kg.

Endoluminal duodenal sleeve

This endoscopically implanted device shields the duodenum and upper jejunum from contact with chyme, thereby mimicking the foregut bypass effect of a gastric bypass procedure without altering the patient’s anatomy. Pancreatic and biliary secretions pass along the outside of the devices and then mix with chyme in the upper jejunum. The device is temporary and needs to be removed after 6–12 months. Based on animal experiments and clinical observations, this device may provide a useful non-surgical intervention for treating type 2 diabetes with an added benefit of providing some weight loss. This device is currently under investigation in industry-sponsored clinical trials.

Gastric pacing and VBloc

Gastric pacing (gastric electrical stimulation) is being experimented with following the work of Cigaina. It consists of a small battery-powered ‘pacemaker’ implanted under the skin beneath the ribs. This device is then attached to two electrodes positioned in the muscle of the anterior gastric wall at the lesser curvature (via laparoscopic surgery). Although the mechanism of action remains unknown, the application of a small electrical impulse to the stomach is thought to induce early satiety and reduce appetite. Another system known as VBloc therapy involves placing the electrodes around the vagal trunks at the level of the gastro-oesophageal junction. Results using these techniques have only demonstrated limited weight loss. Further development of this technology is felt necessary before the device is accepted as bona fide treatment for excess weight.

Clinical effectiveness of surgical intervention

For simplicity, these can be categorised according to the ‘A–O’ Kings criteria.

Airway and apnoea

Severe obesity can be associated with significant respiratory problems. The most important among them are obstructive sleep apnoea and asthma. Among severely obese patients presenting

Table 28.4 Summary of common bariatric operations

	Gastric band	Gastric bypass	Sleeve gastrectomy
Type of surgery	Gastric	Combined	Gastric
Surgical duration	30–45 minutes	60–90 minutes	45–60 minutes
Hospital stay	Day case	2 days	1–2 days
Weight loss (1 year)	20%	30%	25%
Mortality rate	0.01%	0.16%	0.02%
Advantages	Reversible and adjustable Low early operative complication rate	Quick, dramatic weight loss Long clinical track record Good for sweet eaters as ‘dumping’ discourages high-sugar foods	No intestinal bypass Good weight loss
Disadvantages	Use of prosthetic device increases risk of malfunction or infection Risk of band slippage or erosion means 10% reoperation rate long term Regular follow-up – critical for weight loss	Early complications (2%) may be serious 1% long-term risk of bowel obstruction	Early complications (1%) may be serious No long-term data

for interventions such as bariatric surgery more than 70% meet the criteria for obstructive sleep apnoea.

In 2004, a systematic review and meta-analysis reported improvements of sleep apnoea in 85% and resolution in 83% after bariatric surgery. Numerous studies demonstrate an association between obesity and asthma. BMI has a strong, independent and positive relation with asthma. Severely obese patients with asthma experience resolution or improvement after bariatric surgery. A large single-centre study found that 26% of severely obese patients suffer from asthma preoperatively and the resolution after RYGB was 66%. Several groups have reported major reduction in asthma severity after bariatric surgery-induced weight loss.

Body mass index and waist circumference

The average weight loss after medical interventions is between 5 and 10 kg, and for bariatric surgery it is 20–40 kg. Maintenance of weight loss after medical interventions is however poor, while in the SOS study the mean weight loss following surgery over 10 years was 19.9 kg. Mean weight loss is maximal after 1–2 years, after which weight slowly increases for a decade before it stabilizes. After a period of 15 years in the SOS study, patients who had undergone laparoscopic gastric banding and Roux-en-Y gastric bypass had lost 13% and 27% of their total body weight respectively.

Few randomized controlled trials have compared weight loss between surgical procedures. In one of them sleeve gastrectomy seems to achieve the same weight loss as RYGB at 1 year follow-up. Furthermore, Himpens *et al.* compared laparoscopic sleeve gastrectomy and LAGB at 1 and 3 years' follow-up and found that sleeve gastrectomy has a significantly greater weight loss and reduction in BMI. On the other hand, Angrisani *et al.* showed in a randomized controlled trial that the mean weight was significantly better in the LRYGB than the LAGB after 5 years of follow-up.

Cardiovascular disease

Severe obesity increases the frequency and the severity of the metabolic syndrome, which is the major risk factor for cardiovascular disease. Its prevalence in obese patients was nearly 10-fold higher than in non-obese patients. The definition of metabolic syndrome includes the coexistence of any three of the following five features: central obesity, high serum triglyceride levels, low serum high-density lipoprotein cholesterol levels, hypertension and elevated fasting blood glucose levels. The different components of the metabolic syndrome can be addressed with effective pharmacological treatments, but it appears that reduction of visceral adiposity reduces all the components simultaneously. Buchwald reported a 62% improvement of hypertension, 70% of hyperlipidaemia and 86% resolution of diabetes in obese patients undergoing surgery. Interestingly Ahmed *et al.* (2009) have observed an early reduction in blood pressure just a week following LRYGB, suggesting alternative weight independent mechanisms for the change observed.

Bariatric surgery also delays the progression of atherosclerotic disease. In a controlled 4 year interventional study of carotid artery atherosclerosis, obese patients treated with surgery had a threefold lower rate of progression than controls. It is noteworthy that two large retrospective studies show reduced long-term mortality of cardiovascular diseases. Adams found that the cause-specific mortality in the surgery group decreased by 56% for coronary artery disease and Christou reports that the risk of cardiovascular disease decreased by 72%. Moreover, the severity and the duration of obesity are associated with changes in cardiac structure and function. The most frequent abnormalities are the increased cardiac mass, left ventricular hypertrophy and abnormal measures of diastolic left ventricular function. Bariatric surgery has been demonstrated as a successful modality to improve left ventricular hypertrophy as well as systolic and diastolic performance.

Diabetes

Obesity and type 2 diabetes mellitus are likely to be some of the greatest public health problems in the world. There is a strong relationship between the two and the term 'diabesity' has been coined to suggest an overlap of the problems. Approximately half of those diagnosed with type 2 diabetes mellitus are obese, and 20–30% of those who receive bariatric surgery are diabetic.

Medical advances have delivered significant improvements in glycaemic control and the reduction in micro- and macrovascular disease. Remission of type 2 diabetes remains difficult to achieve without the use of bariatric surgery. A recent meta-analysis demonstrates that bariatric surgery has impressive results in the treatment of diabetes, with 78% of diabetic patients demonstrating complete remission and 86% an improvement or remission of diabetes after bariatric surgery. Further randomized controlled trial data from the STAMPEDE trial and Mingrone *et al.* add further weight to the evidence supporting surgery as a treatment option for type 2 diabetes. Regarding the effectiveness of each procedure, diabetes remission is greatest for patients undergoing BPD/DS, followed by gastric bypass, and least for gastric banding. In a randomized controlled trial, Dixon and colleagues compared the effectiveness of laparoscopic gastric banding and conventional therapy in remission of type 2 diabetes mellitus at 2 year follow-up and found that it was significantly higher (73%) in the LAGB group than in the conventional group (13%). Generally, bypass procedures produce the highest and most rapid rates of type 2 diabetes mellitus remission; this is thought to occur by a weight loss-independent mechanism that may involve the role of gut hormones through the so-called enteroinsular axis. Adams in a large cohort study described a remarkable 92% decrease in diabetes-related deaths after RYGB.

Economic complications

The major economic complications for obese patients are unemployment, reduced productivity and the cost of care due to the comorbidities which are related to obesity. The social stigma associated with being overweight may limit a person's ability to get a job. The majority of studies investigating occupational status showed an increase in employment after surgery. Productivity has also been shown to increase significantly after bariatric surgery, with fewer days on sick leave. The SOS trial showed that the number of days lost due to sick leave before the operation was twice as high and disability pension twice as frequent as in the general Swedish population. During the first year after treatment, the surgically treated group had 35% more days of sick leave than the controls, but significantly fewer days during the second and third year. During years 4 and 5, the difference in sick leave between groups was not significant.

In a study on the economic impact of bariatric surgery, the costs of open surgery are fully recovered after 4 years, and for laparoscopic bariatric surgery are fully recovered after 2 years. The cost reductions observed in this analysis mirror the comorbidity reductions in diseases associated with obesity in terms of prescription drug use, hospital visits and physician visits.

Functional (physical functioning)

Severely obese individuals experience significant impairment in activities of daily living such as walking, climbing stairs and bathing. These are reported to be the most distressing aspects of their obesity. Furthermore, obese and overweight individuals face an increased risk of musculoskeletal pain and osteoarthritis. O'Brien in a randomized clinical trial found that physical functioning scores of obese patients who underwent LAGB improved significantly more than medically treated patients. Moreover, the SOS study reports that patients who underwent bariatric surgery reported significant improvement (or resolution) in types of 'work-restricting' pain in the neck, back, hips, knees and ankles. In a prospective study McGoey found that after a mean surgical weight loss of 44 kg the prevalence of back pain decreased from 62% to 11%, hip pain from 11% to 2%, knee pain from 57% to 14%, ankle pain from 34% to 2% and foot pain from 21% to 1%.

Gonadal

Obesity is associated with infertility. In a large, prospective, multicentre study, 42% of women prior to undergoing bariatric surgery experienced subfertility, but 61% had a live birth after surgery. The mechanisms contributing to subfertility in this cohort may include androgen excess, insulin resistance, and hyperinsulinism. Numerous studies found evidence of improvement in fertility after bariatric surgery via the normalization of hormones and menstrual cycles and the reduction of PCOS. PCOS is a syndrome characterized by infertility, menstrual dysfunction, hyperandrogenaemia and hyperinsulinaemia. In the LABS study PCOS was diagnosed in 13% of women undergoing bariatric surgery, a rate significantly higher than that of the general population. In a prospective study of 12 women with PCOS after bariatric surgery all had improvements in hirsutism, hyperandrogenaemia, insulin resistance and restoration of regular menstrual cycles and/or ovulation. In another retrospective study of 24 women who had PCOS and gastric bypass, all had complete resolution of their menstrual irregularity and those who wanted to fall pregnant were able to do so without the need for clomiphene, while 75% had marked improvement in a hirsutism score.

A number of studies have shown that maternal obesity introduces multiple risks for the mother and the fetus during pregnancy. According to the American College of Obstetrics and Gynecology, all bariatric surgery patients are advised to delay pregnancy during the rapid weight loss phase within the first 12–18 months postoperatively. A systematic review of maternal and neonatal complications suggests that the risk for maternal complications, such as gestational diabetes and pre-eclampsia, may be lower following surgically induced weight loss than the risk in obese women and may approach non-obese population rates. Similarly, neonatal complications, such as premature delivery and low birth weight, may be lower in pregnancies following bariatric surgery.

Obesity is also associated with high risk of sexual dysfunction. Many patients present for bariatric surgery with the expectation that weight loss will improve their sexual relations with their partner, and most patients indicate a significant improvement and a more satisfying sexual life postoperatively.

Health status

Previous research has demonstrated that 25–30% of patients had marked clinical symptoms of depression before surgery. The SOS provides the best available evaluation of changes in overall mood, depression and anxiety after bariatric surgery. Poor general health perceptions have frequently been associated with limitations in functional ability, physical and mental symptoms and a number of medical diagnoses. In the SOS study, health perceptions improved by 11% after 10 years. The significant improvement from baseline to 10 year follow-up is in line with the positive long-term effects of weight reduction but was significantly lower than the population norm.

Overall, significant improvements were observed in the surgical group during the first year after bariatric operation. However, the effect on overall mood after 10 years was positive in patients with weight losses of 10% or more, whereas no improvement was observed in patients who lost less than 10% of their initial weight. The depression scores decreased to about 27% at 10 year follow-up. Depression was improved at 10 year follow-up, but the prevalence was still higher than the population norm. A substantial short-term reduction of anxiety symptoms (23%) was seen in the surgical group 10 years after surgery.

Severe obesity is associated with multiple forms of negative health impact that affect quality of life. The SOS study demonstrated that long-lasting weight reduction after bariatric surgery has a general longstanding positive outcome on health-related quality of life. In another recent comparison of gastric bypass surgery with severely obese patients who did not undergo surgery, gastric bypass led to improved health-related quality of life.

Image

Commonly, overweight and obese individuals suffer from body image dysphoria. Several studies have reported the association between weight loss after bariatric surgery and improvement in body image. Adami and colleagues reported that, 3 years after the operation, the scores of patients on the body image dissatisfaction subscale dramatically improved and were similar to the scores observed in normal weight individuals. The great majority of studies indicated a considerable improvement of self-esteem after surgery.

Eating disorders are common in patients undergoing bariatric surgery. Approximately 10–25% of patients meet the criteria for binge-eating disorder (BED) and the prevalence of night-eating syndrome (NES) is between 5% and 20%. BED involves repeated uncontrolled episodes during which objectively large amounts of food are consumed, in association with marked emotional disturbance. Several studies have investigated the relationship between the presence of BED before surgery and the postoperative weight loss, but the results are contradictory. There is evidence that RYGB may improve eating disordered behaviour. Hsu and colleagues report that none of the patients who had BED preoperatively reported BED after RYGB. In another study of LAGB, preoperative BED, uncontrolled eating and NES occurred in 14%, 31% and 17.1% of subjects, respectively. These eating disorders were reduced after surgery to 3.1%, 22.5% and 7.8%, respectively. Having active BED when having surgery may be a poor prognostic feature and hence treatment should be considered before the operation.

Gastro-oesophageal junction

Gastro-oesophageal reflux disease (GORD) has been shown to have a strong association with BMI. Central adiposity may be the most important risk factor for the development of reflux and related complications such as Barrett's oesophagus and oesophageal adenocarcinoma. RYGB seems to be a beneficial surgical treatment for GORD in obese patients as the small stomach pouch reduces acid production and limits reflux. In a prospective study, GORD symptoms decreased in 94% of patients at 9 months postoperatively. Similarly in another study the resolution of GORD after RYGB was 87.6%. Frezza *et al.* examined prospectively the effect of RYGB surgery in 152 severely obese patients with chronic GORD. Approximately 80% of patients had GORD symptom resolution at 6 months follow-up. LAGB is another option for the treatment of obesity and GORD symptoms. Dixon *et al.* demonstrated resolution of all reflux symptoms in 76% and improvement in another 14% of patients undergoing LAGB.

Despite the proposed benefit of banding in reflux, other studies have shown contradictory results and worsening of reflux symptoms after LAGB. A recent systematic review revealed that although LAGB often leads to improvement or even resolution of reflux symptoms in the short term, worsening or newly developed reflux symptoms and oesophagitis were reported in a subset of patients when longer follow-up was available. Further well-designed studies with long follow-up are needed to establish the effect of banding on this disease.

Csendes reports that, in obese patients, the incidence of Barrett's oesophagitis preoperatively was 2%. After RYGB, symptoms of reflux disease, signs of erosive oesophagitis and peptic ulcer disease were no longer present. Additionally, Houghton in a study of five patients with Barrett's oesophagitis preoperatively reports complete or partial regression of Barrett's oesophagitis after RYGB in four of them and improvement in reflux symptoms in all.

Kidney

Obesity is associated with impaired renal parameters, obesity-related glomerulopathy and chronic kidney disease (CKD). Obese patients in the Framingham study who were initially free of CKD were likely to reduce their glomerular filtration rate (GFR) over time. In a study of 61 patients, renal parameters and renal function were markedly improved 24 months after bariatric surgery. The earliest marker of CKD risk is microalbuminuria and is associated with risks of progression to end-stage CKD. A recent meta-analysis found that weight loss is associated with decreased proteinuria and microalbuminuria. The literature is incomplete, but a case series describes improvements or stability of CKD and improved outcomes in renal transplant patients. In a retrospective study in 25 patients with chronic renal disease, 12 months after receiving bariatric surgery the mean GFR increased significantly compared with preoperatively.

Liver

Non-alcoholic fatty liver disease (NAFLD) includes a wide spectrum of liver disease from hepatic steatosis, non-alcoholic steatohepatitis (NASH), fibrosis and cirrhosis. In a review of obese patients undergoing bariatric surgery, the prevalence of steatosis was 91%, the prevalence of NASH was 37% and the prevalence of unexpected cirrhosis was 1.7%. Obesity and insulin resistance are considered to be the main causative factors of NAFLD. Several studies on severely obese patients with NAFLD have shown an improvement in liver disease after bariatric surgery. Dixon studied 36 severely obese patients with NAFLD who underwent LAGB, comparing liver biopsies at a mean 2 year follow-up and showing significant improvements in steatosis, fibrosis and necroinflammatory scores. In a meta-analysis on the effect of bariatric surgery on NAFLD, 92% of patients showed improvement or resolution of steatosis; for steatohepatitis the percentage was 81%, for fibrosis it was 65% and for complete resolution of NASH it was 69%.

Medication

Obesity and its comorbidities are associated with increased prescription drug use. In the SOS study, 52% of the obese individuals were taking medications compared with only 36% of the randomly selected reference population. The obese patients in this study were more often taking medication for cardiovascular disease, pain, psychiatric disorders, diabetes mellitus and asthma. In the SOS study, of the obese individuals in the surgical group taking any medications at baseline, 77% were still taking medications 6 years after the operation.

In RYGB, a large study reports that the average use of medications reduced from 4.4 per patient preoperatively to 1.3 after the procedure. Similarly, in another study, the mean number of medications for the treatment of hypertension, hyperlipidaemia, GORD and type 2 diabetes mellitus in 77 patients decreased from 2.4 preoperatively to 0.2 medications per patient 12 months after RYGB.

Nutrition

Bariatric surgery is related to changes in eating behaviour. Olbers *et al.* showed in a randomized controlled trial that patients after RYGB have a decreased preference for sugary and fatty foods, whereas they increase their preference for vegetables. The SOS study also reports dramatic changes in eating behaviour in the surgical group 6 months after the operation and also at 2 year follow-up.

Other obese comorbidities

Cancer incidence is increased in obese individuals. A recent meta-analysis showed that high BMI is associated with an increased incidence of many types of cancers. Certain recent studies support that bariatric surgery is associated with a reduction in the overall cancer incidence. In the prospective SOS cohort study the number of first-time cancers after inclusion was lower in the surgery group than in the control group. Moreover, the SOS study concluded that bariatric surgery had a significant effect on cancer incidence in women but not in men.

In two retrospective cohort studies of obese patients treated with bariatric surgery there was also a reduction in cancer incidence in patients who underwent bariatric surgery compared with those who did not. Adams found that mortality in the surgery group decreased by 60% for cancer during a mean follow-up of 7.1 years and Christou, after a follow-up of 5 years, found that the post-bariatric surgery group have significantly fewer visits to the physician/hospital for 'all cancer' diagnosis than the non-operated, severely obese control group. There were only 2% reported bariatric surgery patients with cancer compared with 8.5% control subjects. A recent retrospective study focusing on breast and endometrial cancer incidence showed that bariatric surgery may decrease cancer development.

GUIDE TO FURTHER READING

Ahmed AR, Rickards G, Coniglio D, *et al.* Laparoscopic Roux-en-Y gastric bypass and its early effect on blood pressure. *Obes Surg* 2009;**19**: 845–849.

Buchwald H, Avidor Y, Braunwald E, *et al.* Bariatric surgery: a systematic review and meta-analysis. *JAMA* 2004;**292**:1724–37. Erratum in *JAMA* 2005;**293**:1728.

Buchwald H, Estok R, Fahrbach K, *et al.* Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. *Am J Med* 2009;**122**:248–256.

Christou NV, Sampalis JS, Liberman S, *et al.* Surgery decreases long-term mortality, morbidity, and health care use in morbidly obese patients. *Ann Surg* 2004;**240**:416–424.

Dixon AF, Dixon JB, O'Brien PE. Laparoscopic adjustable gastric banding induces prolonged satiety: a randomized blind crossover study. *J Clin Endocrinol Metab* 2005;**90**:813–819

Longitudinal Assessment of Bariatric Surgery (LABS) Consortium, Flum DR, Belle SH, *et al.* Perioperative safety in the longitudinal assessment of bariatric surgery. *N Engl J Med* 2009;**361**:445–454.

Mingrone G, Panunzi S, De Gaetano A, *et al.* Bariatric surgery versus conventional medical therapy for type 2 diabetes. *N Engl J Med* 2012;**366**:1577–85.

Pournaras DJ, Osborne A, Hawkins SC, *et al.* Remission of type 2 diabetes after gastric bypass and banding: mechanisms and 2 year outcomes. *Ann Surg* 2010;**252**:966–971.

Schauer PR, Kashyap SR, Wolski K, *et al.* Bariatric surgery versus intensive medical therapy in obese patients with diabetes (The STAMPEDE trial). *N Engl J Med* 2012;**366**:1567–76.

Sjöström L, Narbro K, Sjöström CD, *et al.* Swedish Obese Subjects Study. Effects of bariatric surgery on mortality in Swedish obese subjects. *N Engl J Med* 2007;**357**:741–752.

Sjöström L, Gummesson A, Sjöström CD, *et al.* Swedish Obese Subjects Study. Effects of bariatric surgery on cancer incidence in obese patients in Sweden (Swedish Obese Subjects Study): a prospective, controlled intervention trial. *Lancet Oncol* 2009;**10**:653–662

Wyles SM, Ahmed AR. Tips and tricks in bariatric surgical procedures: a review article. *Minerva Chir* 2009;**64**:253–64.

CHAPTER 29

Disorders of the small intestine and vermiform appendix

SIR ALFRED CUSCHIERI

Anatomy

The small intestine is divided into three anatomical portions: duodenum, jejunum and ileum. However, in surgery, the duodenum, which forms the first 25 cm, is considered with the stomach because of various pathologies which affect both organs, such as peptic ulceration. At postmortem the average length of the small intestine *in situ* is about 7 m, with a reported range of 4.3–10.3 m when the measurement is made along the antimesenteric border. These estimates have to be regarded as approximate and are certainly not exact in view of the distensibility of the small intestine, the flaccid state after death and indeed the method of measurement (usually by tape) used in these postmortem studies. The length of the intestine (less the duodenum) when measured along its attached or mesenteric border, is only 4.6 m.

The tone of the smooth muscle reduces the small bowel length considerably during life. Accurate assessment of intestinal length at operation is thus difficult owing to the changing state of the intestinal loops due to peristaltic activity, handling and exposure. In the adult, the length of the small bowel measured along its antimesenteric border and in the unstretched state after a preliminary laparotomy averages 3.5 m. After substantial resections of the small bowel, it is important to measure and record the extent of the residual small intestine in the operation, as this is more useful in subsequent management and prognosis than the extent of resected bowel.

Small bowel mesentery

The small bowel mesentery (SBM) is a fat-laden peritoneal reflection which attaches the jejunum and ileum to the posterior abdominal wall. This attachment, known as the root of the small bowel mesentery (RSBM) and measuring 15 cm, runs obliquely down from the ligament of Treitz to the ileocaecal region and overlies a bare area continuous with the anterior pararenal space of the retroperitoneum. The RSBM contains two major vessels, the superior mesenteric artery and vein (SMA, SMV), in addition to other important structures. Superiorly, the RSBM is continuous with the hepatoduodenal ligament around the SMV and portal vein. Anteriorly, it is continuous with the transverse mesocolon, and posterolaterally with the ascending and descending mesocolon and anterior pararenal space. The right gastrocolic trunk marks the junction between the transverse mesocolon and the RSBM. The inferior mesenteric vein joins the SMV or splenic vein on the left side of the RSBM.

The length (depth in the supine position) of the SBM is important surgically, especially in the creation of Roux-en-Y loops. It increases gradually from the duodenojejunal flexure (ligament of Treitz) and reaches its maximum at 1–1.25 m, where it measures 18–20 cm and then diminishes in length, such that at the mid-ileum it measures 12–13 cm and then shortens rapidly as it approaches the caecum. The lower part of the SBM forms a fold that can be easily palpated during surgery. It is a useful landmark for access to the fossa on the left side of the mesentery for irrigation in patients with generalized peritonitis.

Jejunum and ileum

The jejunum, which constitutes the proximal two-fifths, and is about 2.5 m long, for the most part occupies the umbilical and left iliac regions, whereas the ileum occupies chiefly the umbilical, hypogastric, right iliac and pelvic regions; the distal ileum tends to be situated in the right iliac fossa and pelvis. The terminal ileum usually lies in the pelvis, from which it ascends over the right psoas muscle and right iliac vessels to the right iliac fossa to open into the medial side of the caecum. There is no clear demarcation line between the jejunum and the ileum, but the proximal jejunum is thicker and more vascular than the ileum, has a darker colour with prominent plicae circulares and has an overall diameter of 4.0 cm. The circular folds of its mucous membrane are large and thickly set (thus palpable between finger and thumb as distinct from the ileum), and its villi are larger than in the ileum. The aggregated lymph nodules are almost absent in the upper jejunum, and are sparse, smaller and tend to be circular compared with the ileal Peyer's patches. The mesentery of the jejunum contains less fat and the mesenteric vasculature consists of prominent arteries and veins which join to form one or two arcades in the mesentery before giving rise to the terminal intestinal branches.

The ileum is narrower (average diameter 3.0–3.5 cm), thinner and less vascular than the jejunum. It possesses but few circular folds which are small (impalpable externally) and which disappear entirely in its terminal reaches. In contrast, the ileum contains larger and more numerous aggregated lymph nodules (Peyer's patches). The mesenteric vasculature of the ileum is more complex, with the vessels forming four or five levels of arcades before the origin of the terminal intestinal branches.

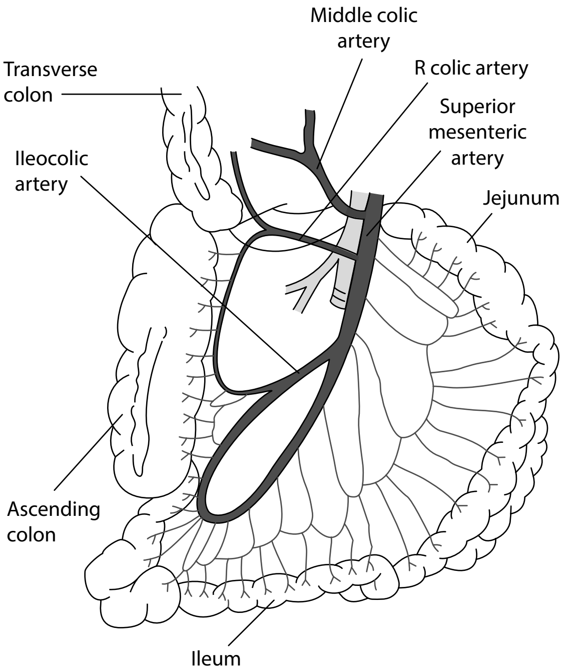


Figure 29.1 Superior mesenteric artery and vein.

Vascular supply

The whole of the small bowel is supplied by the SMA (Figure 29.1). An understanding of the anatomy of the mesenteric vascular arcades of the small intestine is essential for the safe execution of reconstructive procedures on the gastrointestinal tract involving small bowel segments, e.g. ileal pouch construction.

The intestinal branches of the SMA on reaching the attached border of the bowel run between the serous and muscular coats with frequent crossing over of the free border of the intestine, where they anastomose with other branches running around the opposite surface of the small bowel. These vessels give off numerous branches, which pierce the muscular coat, supplying it and forming an intricate plexus in the submucous layer from which minute vessels supply the glands and villi of the mucous membrane. The veins run a similar course and arrangement to the arteries. The lymphatics of the small intestine (lacteals) are arranged in two sets: those draining the mucous membrane and those of the muscular coat. The lymphatics of the intestinal villi form an intricate plexus in the mucous and submucous layer, where they are joined by the lymphatics from the lymph spaces at the bases of the solitary nodules. Thereafter they proceed as larger channels (lacteals) at the mesenteric border of the small intestine. The lymphatics of the muscular coat are situated between the two muscular coats, where they form a plexus, but throughout their course they communicate freely with the lymphatics from the mucous membrane, and ultimately empty in the lacteal vessels.

The arteriolar supply to the small bowel is of particular surgical interest. Each arteriole and accompanying venule supply and drain one-half of the circumference of the small intestine, and the vessels are distributed to alternate sides of the bowel wall in regular sequence (Figure 29.2). This arrangement forms the basis of intestinal lengthening operations in infants and children with short gut syndrome as the residual bowel can be split longitudinally, with each half retaining an adequate blood supply (see section Short gut syndrome).

Microanatomy

The small bowel mucosa is arranged in villi and crypts. The villi form finger-like projections measuring 0.5–1.5 mm in length (Figure 29.3a,b), with the crypts forming depressions below the surface in between the projecting villi. These are

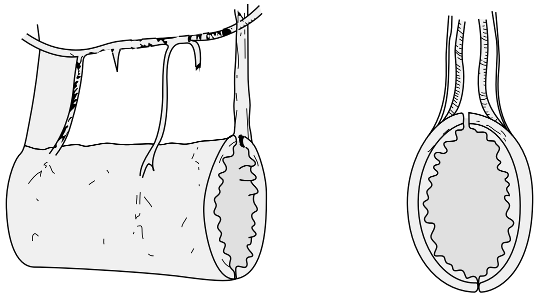


Figure 29.2 Diagrammatic representation of the distribution of the terminal branches of the mesenteric vessels to the small intestine. These are distributed to alternate sides of the gut.

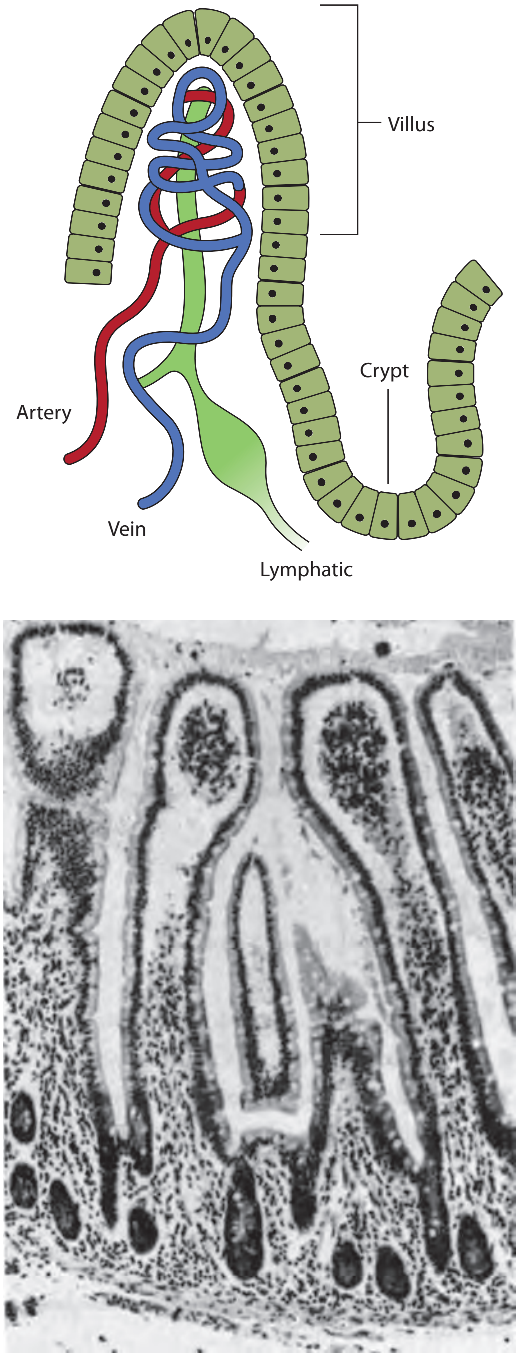


Figure 29.3 (a) Schematic anatomy of the small bowel villus and crypt. (b) Microphotograph showing normal appearance of jejunal mucosa.

covered by tall columnar mature absorptive cells (enterocytes) which have microvilli at their luminal surfaces (brush border). These enterocytes rest on a thin basal membrane (lamina propria), and their microvilli are kept lubricated by a surface mucus known as the glycocalyx, which separates the brush border from the intestinal contents. Goblet cells, found interspersed among the enterocytes, synthesize the mucinous glycoprotein essential for maintaining the glycocalyx. At their bases, the villi are surrounded by intestinal crypts (of Lieberkuhn) which are lined by young epithelial secreting cells. Stem cells are found at the base of the crypts. These are constantly dividing in order to replace the surface epithelium by a process of cell division and migration. The divisions of these stem cells produce new stem cells (which remain pluripotential) and other cells which are committed to differentiate along one of four lineages to become enterocytes, enteroendocrine, goblet and Paneth cells. The cells following the enterocyte lineage divide several more times as they migrate up the crypts. During this migration, they differentiate further into the mature absorptive cells that express all the transport proteins and enzymes characteristic of the mature enterocytes.

The other cells found in the crypts are goblet cells in the upper half, Paneth cells at the base and endocrine cells. The Paneth cells are pyramidal in shape and have abundant RNA-rich cytoplasm with refractile granules containing lysosomes. The exact function of Paneth cells is unknown, but they are thought to be capable of phagocytosing bacteria from the crypt lumen. The endocrine cells are also known as APUD cells. This term is derived from a basic characteristic of these enteroendocrine cells: uptake and decarboxylation of amine precursors. These gut endocrine cells were once thought to be of ectodermal origin but are now thought to arise from the undifferentiated crypt stem cells. A separate group of gut endocrine cells stain with potassium chromate and silver dyes. These are known as enterochromaffin cells or argentaffin cells. These synthesize and secrete 5-hydroxytryptophan (5-HTP) and 5-hydroxytryptamine (5-HT) in addition to a variety of peptide hormones such as motilin, substance P and others. The other important cellular components of the villus are lymphocytes, which are subdivided into those in the basal lamina (lamina propria lymphocytes) and those found among the epithelial cells covering the villus (intraepithelial lymphocytes).

Physiology

The primary function of the small intestine is the digestion and absorption of nutrients. The other functions of the small bowel include synthesis of lipoproteins and various hormones.

Transport

Transport of molecules and ions across the epithelium of the small intestine occurs by two routes: (1) the transcelluar route across the plasma membrane of the epithelial cells and (2) the paracellular route across tight junctions between the epithelial cells with some molecules, e.g. water, being transported by both routes. As the tight junctions are impermeable to large dietary organic molecules, e.g. amino acids and glucose, these are transported exclusively by the transcellular route through the enterocytes which have the necessary specific transporter molecules.

The absorptive epithelium of the gut is not uniform throughout its length. In particular, the permeability of the tight junctions linking cells varies considerably in the different regions of the small intestine. As the digested food travels caudally, it is exposed sequentially to regions of mucosa with different absorptive characteristics as a result of phenotypic differences of the enterocytes, i.e. the number and type of transporter molecules expressed and the permeability of the tight junctions. Even within a given segment of the small intestine, there are major differences in the types of transport, e.g. the transport in crypt cells differs from that of the villi in any given segment.

Small bowel secretion

Large quantities of water are secreted into the lumen of the small intestine during the digestive process. Almost all of this water is reabsorbed in the small intestine. Regardless of whether it is being secreted or absorbed, water flows across the mucosa in response to osmotic gradients. In the case of secretion, two distinct processes establish an osmotic gradient that pulls water into the lumen of the intestine.

1. *Increases in luminal osmotic pressure resulting from entry and digestion of chyme:* on entry into the small bowel, chyme is further digested in the small bowel with a substantial increase in the osmolarity. As the small bowel luminal osmolarity increases with digestion of foodstuffs, a net transfer of water into the lumen of the small bowel occurs. Thereafter, as the osmotically active molecules (maltose, glucose and amino acids) derived from digestion are absorbed, the osmolality of the intestinal contents decreases enabling the water to be reabsorbed.
2. *Active secretion of electrolytes by crypt cells induces water secretio*n: the luminal membrane of crypt epithelial cells contains the ion channel: cyclic adenosine monophosphate (cAMP)-dependent chloride channel, also known as the cystic fibrosis transmembrane conductance regulator (CFTR). Mutations in the gene for CFTR are the cause of cystic fibrosis. CTFR enables secretion of water as follows.
   - Cl^- enters the crypt epithelial cell by cotransport with Na^+ and K^+, but Na^+ is removed from the cell by sodium pumps, and K^+ is exported through other channels.
   - Activation of adenyl cyclase by secretagogues with generation of cAMP.
   - Raised intracellular concentration of cAMP in crypt cells activates the CFTR, resulting in secretion of chloride ions into the lumen.
   - Accumulation of Cl^- in the crypts creates an electric potential attracting Na^+ into the small bowel lumen across tight junctions – hence the secretion of NaCl.
   - Finally, the NaCl in the crypt creates an osmotic gradient across the tight junctions, drawing water into the lumen.

Abnormal activation of the cAMP-dependent chloride channel (CFTR) in crypt cells from bacterial toxins which activate the adenylate cyclase in crypt enterocytes is responsible for the abnormal massive secretion of water, with severe dehydration and not infrequent death. Cholera toxin, produced by *Vibrio cholerae,* is the best known example, but several other bacterial species can induce the same potentially lethal pathology.

Small bowel absorption

It is important to stress that the single most important process occurring in the small bowel, which underlies absorption, is the establishment and maintenance of an electrochemical gradient of sodium across the epithelial cell boundary of the small bowel lumen. To remain viable, all cells have to maintain a low intracellular concentration of sodium. In the enterocytes, low intracellular sodium is maintained by a large number of Na^+/K^+ ATPases (sodium pumps) embedded in the basolateral membrane. These pumps export three sodium ions from the cell in exchange for two potassium ions, thus establishing a gradient of both charge and sodium concentration across the basolateral membrane (Figure 29.4). In essence, it is the flow and accumulation of sodium that is ultimately responsible for absorption of water, amino acids and carbohydrates.

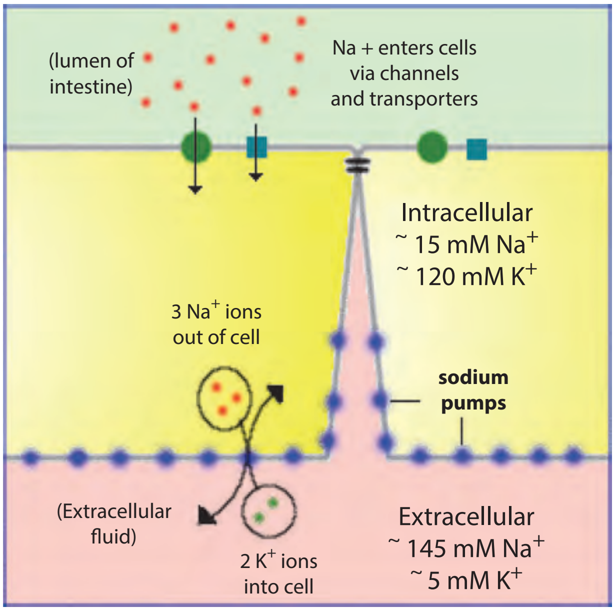


Figure 29.4 Schematic diagram illustrating the establishment of the electrochemical gradient across through the activity of sodium pumps.

Absorption of monosaccharides

Monosaccharides are derived by enzymatic digestion of more complex carbohydrates (polysaccharides). Particularly important dietary carbohydrates in man include starch and disaccharides, especially lactose and sucrose. All these molecules cannot be absorbed without prior luminal enzymatic digestion into monosaccharides by enzymes present in the brush border of the luminal plasma membrane of absorptive enterocytes and known collectively as *brush border hydrolyases*. These include maltase, lactase and sucrase. Whereas dietary lactose and sucrose are directly digested by their respective brush border enzymes, starch requires preliminary digestion to maltose by salivary and pancreatic amylases. At the brush border, maltase splits maltose into two molecules of glucose, lactase splits lactose into galactose and glucose and sucrase splits sucrose into glucose and fructose. The resultant glucose and galactose enter the enterocyte by cotransport with sodium using the same transporter, whereas fructose uses a separate transporter (GLUTS) for its entry into the enterocyte.

Transport of glucose and other monosaccharides across the intestinal epithelium

The transporter that carries glucose and galactose into the enterocyte is the sodium-dependent hexose transporter, known as SGLUT-1, which transports both glucose and sodium ions into the cell through a series of conformational changes as follows.

1. The transporter facing the lumen initially binds sodium.
2. In turn, the bound sodium induces a conformational change which opens the glucose-binding pocket of the transporter.
3. As glucose binds to the transporter, this reorientates such that the pockets carrying sodium and glucose are moved intracellularly.
4. Sodium then dissociates into the cytoplasm, thereby destabilizing the glucose binding when glucose dissociates from the transporter into the cytoplasm.
5. The unloaded transporter reorientates back to its original luminal-facing position.

Once inside the enterocyte, glucose and sodium are transferred into the bloodstream. Sodium is rapidly exported in exchange for potassium by the sodium pumps, thereby maintaining the electrochemical gradient across the epithelium. Glucose, galactose and fructose are transported out of the enterocyte into the capillary bloodstream of the villus by means of another hexose transporter (called GLUT-2) also located in the basolateral membrane.

Absorption of amino acids and peptides

Dietary proteins have to be digested prior to absorption. The enzymes involved are gastric pepsin and a group of pancreatic proteases, also secreted as zymogens and requiring activation to trypsin, chymotrypsin and carboxypeptidases. By these enzymes, dietary proteins are hydrolysed within the lumen of the small intestine predominantly into amino acids, medium and small peptides (oligopeptides). The brush border peptidases are responsible for the further hydrolysis of luminal peptides into free amino acids and very small peptides. The di- and tripeptides, which have a high affinity for the brush border peptidases, are hydrolysed to individual amino acids before absorption. Those with a low affinity for the brush border enzymes are absorbed intact by the enterocytes.

The luminal plasma membrane of the enterocytes also has sodium-dependent amino acid transporters – one each for acidic, basic and neutral amino acids. These transporters bind amino acids only after binding sodium. The fully loaded transporters then undergo a conformational change that releases the sodium and the amino acid into the cytoplasm, followed by its reorientation back to the original form. Thus, as with monosaccharides, the absorption of amino acids is also dependent on the electrochemical gradient of sodium across the epithelium. The basolateral membrane of the enterocyte contains additional transporters which export amino acids from the enterocytes into the bloodstream but these are not dependent on sodium gradients.

Both di- and tripeptides are also absorbed in abundance by the mucosa of the small intestine by cotransport with H^+ ions via a transporter called PepT1. Once inside the enterocyte, the majority of the absorbed di- and tripeptides are digested into amino acids by cytoplasmic peptidases and exported from the cell into blood, with only a very small fraction entering the bloodstream intact. For a few days after birth, neonates have the ability to absorb intact proteins. This ability, which is rapidly lost, is of immense importance because it enables the newborn baby to acquire passive immunity by absorbing immunoglobulins present in colostral milk.

Absorption of lipids

Although the bulk of dietary lipid is made up of triglycerides, food also contains phospholipids, sterols especially cholesterol and many other lipids, including fat-soluble vitamins. Triglyceride absorption entails two processes: (1) emulsification followed by (2) enzymatic digestion to monoglycerides and fatty acids.

Bile acids are responsible for emulsification by virtue of their amphipathic nature, i.e. they have both hydrophilic and hydrophobic domains. On exposure to large aggregates of triglycerides, the hydrophobic portions of bile acids intercalate and surround the lipid aggregates in a specific way that results in the hydrophilic domains being exposed at the surface. Aside from the breaking up of large lipid aggregates into progressively smaller ones, the external hydrophilic domains ensure aqueous dispersion in the fluid of the small bowel content.

Pancreatic lipase and colipase are responsible for the hydrolysis of the emulsified triglycerides by splitting the fatty acids at positions 1 and 3 of the triglyceride chain with the production of two fatty acids and one 2-monoglyceride.

The monoglycerides and fatty acids produced by the action of lipase retain their association with bile acids and complex with other lipids to form micelles, which are essentially small aggregates (4–8 nm in diameter) of mixed lipids and bile acids suspended within the intestinal fluid. At the surface of the enterocytes the monoglycerides and free fatty acids separate from the micelles before entering the enterocyte, leaving the bile acids within the intestinal lumen. At the brush border lipids, including monoglycerides and fatty acids, are taken up into the epithelial cells.

Lipid transport into the blood

Both fatty acids and 2-monoglycerides enter the enterocytes by simple diffusion, but considerable amounts of the fatty acids are transported into the enterocytes via a specific fatty acid transporter protein. Once inside the enterocyte, fatty acids and monoglycerides are bound to fatty acid-binding proteins (FABPs) and transported to the smooth endoplasmic reticulum where they are re-esterified. Within the Golgi apparatus of the endoplasmic reticulum, the triglycerides are converted to chylomicrons consisting of cholesterol and triglyceride surrounded by phospholipids, free cholesterol and apoproteins. These chylomicrons together with lipoproteins are transferred via the lacteals and lymph channels and ultimately through the thoracic and left lymphatic trunks into the venous blood.

The absorption of medium chain triglycerides (MCTs) on the other hand is different. A significant percentage of MCTs is absorbed intact into the portal blood. The rest is broken down by pancreatic lipase to medium chain fatty acids which are readily absorbed by the intestinal cells after micellar aggregation with bile acids. These are then transferred as free fatty acids into the portal venous blood without chylomicron formation.

Cholesterol homeostasis results from a balance of cholesterol synthesis, absorption of dietary cholesterol and excretion in bile. A specific transport protein (NPC1L1) is responsible for transport of cholesterol from the intestinal lumen into the enterocyte, where the bulk is esterified and incorporated into chylomicrons.

Under normal circumstances, the digestion and absorption of fluid, electrolytes, iron, folates, carbohydrates, fat and proteins is completed in the jejunum. However, the effective absorption of bile salts and vitamin B_{12} can only occur in the terminal ileum, as these substances require specific transport sites which are located in this region. In the absence of ileum the critical length of jejunum required for the absorption of water-soluble substances is somewhere between 30 and 120 cm. However, the absence of the ileum will produce malabsorption of fat from a diminished bile salt pool and of vitamin B_{12}. The ileum on the other hand is able to perform all the absorptive functions of the jejunum if the proximal small bowel is excised.

Other functions of the small intestine

Another function of the small intestine is the synthesis of high-density, low-density and very low-density lipoproteins (HDL, LDL, VLDL). These are closely related to the chylomicrons and contain the same apoproteins. The intestinal lipoproteins reach the plasma via the thoracic duct. Although there are other tissues which synthesize these lipoproteins, e.g. the liver, the gastrointestinal tract is a major site of production and therefore plays an important part in the metabolism of plasma lipoproteins. The synthesis of these lipoproteins by the small intestine is impaired in kwashiorkor.

The third function of the small intestine is the synthesis of peptide and amine intestinal hormones. These are located within the enteroendocrine cells and in the neurones of the myenteric plexus. These modulate intestinal activity in three ways: (1) as classic endocrine hormones, (2) neurotransmitters and (3) as hormones with a paracrine action. These hormones influence intestinal secretion and transport, growth and differentiation, the splanchnic haemodynamic state and the release of insulin (Table 29.1).

Table 29.1 Hormones originating in the small bowel and their functions

Hormone	Functions
Acetylcholine	Neurotransmitter
Enkephalins	Neurotransmitter
Enteroglucagon	Growth and differentiation
Gastric inhibitory peptide	Glucose-dependent insulin release
Glucagon	Increased splanchnic blood flow
Motilin	Secretion and transport
Neurotensin	Increased splanchnic blood flow, secretion and transport
Noradrenaline	Neurotransmitter
Secretin	Secretion and transport
Serotonin	Neurotransmitter
Somatostatin	Neurotransmitter, secretion and transport
Substance P	Neurotransmitter
Vasopressin	Decreased splanchnic blood flow
Vasoactive intestinal peptide	Increased splanchnic blood flow

The complex hormonal interactions within the small intestine are closely linked with the activity of the enteric nervous system, which is considered to act as an independent integrative system regulating reflex activity within the gut. Within the myenteric plexus, neurones containing opioid peptides (enkephalins), substance P, vasoactive intestinal peptide (VIP), serotonin and somatostatin have all been identified in addition to the better known adrenergic [noradrenaline (norepinephrine)] and cholinergic (acetylcholine) neurotransmitters.

Small intestinal motility

The co-ordinated contractions of the small intestine after a meal (fed pattern) are designed to facilitate digestion and absorption by:

- ensuring adequate mixing of the ingested foodstuffs with the pancreatic digestive enzymes and with bile salts entering the small bowel through the sphincter of Oddi as a result of co-ordinated contractions of the gallbladder after a meal induced by cholecystokinin
- constant dispersal of nutrient molecules in the bowel lumen derived by digestion, thus bringing them in close contact with the enterocytes where enzymatic digestion is completed and absorption occurs
- advancing the chyme down the small bowel, thereby making space for the next load delivered from the stomach.

Motility in the small intestine is controlled predominantly by excitatory and inhibitory signals from the enteric nervous system modulated by input from the central nervous system (CNS) and a number of gastrointestinal hormones. Studies of gastrointestinal motility are difficult to undertake and to interpret. Most have been performed with multilumen fluid-perfused catheters attached to external transducers linked to chart recorders. However, solid-state systems based on several miniature strain gauge transducers mounted at intervals on a fine catheter are now commonly used. The analogue signals (corresponding to the intestinal contractions) from these transducers are stored in a solid-state external portable logger. The data are then offloaded onto a computer for analysis. Software systems are now available which are capable of sophisticated analysis of small bowel activity.

Using these new systems it has been possible to characterize small bowel motility both in normal and in pathological states. In the normal situation both fasting and fed patterns of intestinal activity are now recognized and the types of intestinal contractions fall into two categories, *mixing* and *propagating*. The mixing contractions include the *stationary ring contractions*, which occlude the lumen at one point and push the intestinal contents in both directions, and the *stationary cluster contractions*, which may occur at one or several sites simultaneously. The propagating contractions include *single propagated contractions*, *propagating power contractions*, *migrating cluster contractions* and *the phase III of the migratory motor complex (MMC)*. The propagating power contractions are also known as the giant migratory contractions. They correspond to the peristaltic rushes described by earlier workers and are effective at sweeping intestinal contents in a distal direction at a rapid rate. They are increased in patients undergoing radiotherapy and are now thought to be the cause of the nausea, vomiting, abdominal pain and diarrhoea which are often experienced by these patients.

In the interdigestive period between meals, the small bowel is empty due to the MMC consisting of contractions which originate in the stomach and sweep down the entire length of the small intestine, sometimes causing audible borborygmi.

Motility disorders may be functional (no definable organic cause) or secondary to myopathy (hollow visceral type), autonomic neuropathy, abnormalities of the enteric nervous system, endocrine disorders, tumours, drug-induced and multiple endocrine neoplasia. The motility disorders of the small and large intestine that are of surgical importance include adynamic (paralytic) ileus, irritable bowel syndrome, slow transit constipation and Hirschsprung disease.

Microbiology

At birth the alimentary tract is sterile but becomes colonized by bacteria via the oral route, so that within 3–4 weeks of birth the enteric microflora found in the adult are already established. The stomach, duodenum and proximal jejunum in the normal adult subject contain transient Gram-positive aerobes such as lactobacilli and enterococci in concentrations of less than 10^4 colony-forming units (CFU) per mL of contents. In the terminal ileum, however, the bacterial counts are higher (10^5–10^8 CFU) and the flora resemble that of the large bowel. The organisms encountered include coliforms and strict anaerobes (e.g. bacteroides). Substantially higher bacterial counts (10^5–10^{11} CFU) are found in the colon, with bacteroides, anaerobic lactobacilli and clostridia predominating. It is thought that the small bowel microflora play an important protective role against bacterial overgrowth and infection by preventing colonization of the surface epithelium by pathogenic bacteria. This is brought about by production of bacteriocins (antibiotics) and lowering of the oxidation–reduction potential by anaerobic metabolism, which results in the production of substances that are toxic to bacteria.

Intraluminal bacteria produce endotoxin which is not absorbed into the bloodstream in the presence of bile salts. However, absorption of endotoxin into the blood occurs whenever bile salts are prevented from reaching the small intestine as in obstructive jaundice. This endotoxinaemia may be responsible for the development of renal failure after surgical intervention in jaundiced patients. The small bowel microflora play an important role in intraluminal metabolism of various substances, particularly metabolized protein and other nitrogenous compounds, the results of which are absorbed and used for synthesis of amino acids. For example, the excess ammonia produced by bacterial metabolism after an episode of gastrointestinal haemorrhage may precipitate encephalopathy in patients with chronic liver disease and portal hypertension. Under normal conditions a symbiotic relationship exists between the gut and intestinal organisms, and indeed the villous architecture and rate of regeneration of the intestinal mucosa depend on the presence of a normal resident microflora. However, in conditions of stasis and diminished or absent gastric secretory activity, bacterial overgrowth may occur leading to maldigestion and malabsorption.

Gut mucosal integrity and bacterial translocation

The normal intestinal mucosa in the presence of a competent immune system is able to resist invasion by pathogenic bacteria. This antimicrobial barrier can break down under certain pathological states with translocation of the pathogens to the blood and lymph, with subsequent systemic invasion of tissues and organs. Such bacterial translocation underlies one of the hypotheses held responsible for systemic inflammatory response syndrome, i.e. endotoxin-induced activation of the cytokine cascade, which in turn leads to multiorgan failure (MOF) and death. This hypothesis has, however, been challenged recently, because treatment with antibodies to the polysaccharide fraction of endotoxin has been shown to be ineffective in clinical trials, and other mechanisms, e.g. endothelial damage by activated neutrophils, may be responsible for MOF. The conditions that promote bacterial translocation are:

- bacterial overgrowth
- immunodeficiency
  - physical disruption of mucosa including ischaemia
- trauma
  - burns
- sepsis
- endotoxinaemia and protein malnutrition.

In septic patients enteral feeding with complete diets may nonetheless protect the integrity of the gut barrier and improve the immune function and is thus superior to parenteral feeding. The components of enteral feeds which have been shown to enhance the ability of the intestinal mucosa to resist bacterial translocation are arginine, glutamine and lipids (especially fish oil). Arginine, a dibasic amino acid, can stimulate the secretion of insulin, prolactin and growth hormone and also has immunomodulatory effects. Thus in immunocompromised patients, arginine supplements have been shown to enhance lymphocyte function in the postoperative period. There is also good evidence that dietary lipids may have a similar effect. This may act directly or indirectly via alterations in eicosanoid metabolism. The eicosanoids include prostaglandins and leukotrienes that modulate the inflammatory response and immune function. Glutamine is the most abundant amino acid in the free amino acid pool and is essential for all rapidly dividing cells including enterocytes. It thus plays an important role in maintaining the integrity of the intestinal mucosa. Although the glutamine pool is large, it is labile and becomes rapidly depleted in injured and septic patients. Thus in the critically ill, glutamine may be regarded as an essential amino acid. There is good evidence that the translocation of bacteria encountered in sepsis, trauma and burns may be reversed by glutamine-enriched intravenous or enteral nutrition. This positive effect is accompanied by a normalization of the secretory IgA levels and a decrease in the bacterial adherence to enterocytes.

Small bowel investigations

Specific investigation of the small intestine is indicated where small bowel obstruction, bleeding or malabsorption are suspected. The term malabsorption can be used to describe the failure to absorb specific substances, e.g. carbohydrates, fats, proteins, minerals, etc., but when used unqualified it generally refers to fat malabsorption. This is frequently encountered in surgical practice as it may follow excisional and bypass procedures in the gastrointestinal tract which effectively shorten the small bowel and may contribute to bacterial overgrowth. Clearly, serum biochemistry, e.g. albumin, transferrin, electrolytes, including calcium, iron, haemoglobin level and blood film, are all necessary when malabsorption is suspected. There are however special investigations which may assist in establishing the diagnosis.

Radiology

The radiological investigations that are of value in resolving pathologies which may affect the small bowel include plain radiography, barium studies, fistulography, ultrasound and CT scanning, mesenteric angiograph and isotope studies.

Plain radiography

Plain abdominal radiographs provide useful information in the diagnosis of the acute abdomen. As far as the small bowel is concerned, obstruction is diagnosed by visualizing dilated gas-filled loops of small bowel. These can be distinguished from large bowel by the more central distribution and the ladder-like valvulae conniventes. Traditionally, an erect abdominal film is obtained at the same time but this does not provide any extra information, although it is only on the erect film that the typical fluid levels will be seen. Occasionally, however, obstructed small bowel may be so full of fluid that it will not be seen clearly on the supine film and the fluid levels will only be seen on the erect film which will give the diagnosis. Thickening of the small bowel can be inferred when there is a significant gap between the gaseous outline of the lumen of adjacent loops of small bowel. Plain radiography is also very useful for identifying free air in the peritoneum, but it must be remembered that an erect chest radiograph is more useful in this respect, as in the erect abdominal film the diaphragm may not be included.

Barium studies

The mainstay of the investigation of the small bowel is the barium small bowel follow-through which is easy to perform as part of a barium meal study. However, a small bowel enema in which contrast medium is instilled via a Bilbao–Dotter tube directly into the upper jejunum carries a higher diagnostic yield, especially for small bowel tumours and in patients with suspected malabsorption. In particular, ulcers, sinuses and fistulas are better visualized by a small bowel enema (enteroclysis). The radiological criteria of malabsorption are flocculation and segmentation of the barium, thickening of the mucosal folds and dilatation of intestinal loops. It must be stressed however that these changes are non-specific and the diagnosis must be confirmed by other more specific tests.

There have been few comparative studies between barium meal and follow-through and small bowel enema (enteroclysis), but available evidence indicates that the latter is superior except in Crohn's disease. The false-negative rate for the detection of primary small bowel neoplasms is much higher (80%) for follow-through examinations than small bowel enema (10%). Enteroclysis is better than follow-through and even radionuclide studies for the detection of Meckel's diverticulum. Small bowel enema is useful in the diagnosis of partial/intermittent obstruction and is superior to both follow-through and CT in these cases. In contrast, CT is accurate in showing site and cause in patients with established high-grade obstruction.

In the patient with obstruction due to adhesions where there are some doubts as to whether or not the obstruction will settle with conservative treatment, it is useful to give the patient a drink or nasogastric bolus of water-soluble contrast such as 40 mL of Urografin in 40 mL of distilled water. If the contrast has not reached the ascending colon within 24 hours, it is highly likely that the patient will require surgical intervention.

Injection of contrast into the fistula may also be of value (fistulography). This is particularly useful in Crohn's disease. Where enterocutaneous fistulas have formed, fistulography will allow the diagnosis of associated abscess cavities and strictures and will aid the planning of appropriate surgical intervention.

Ultrasound and CT scanning

Ultrasound scanning may be used in patients with intestinal obstruction and is capable of differentiating fluid-filled dilated intestinal loops from other cystic structures within the abdomen, but its role in the investigation of small bowel disease is limited. CT scanning of the small intestine entails the use of a special barium sulphate suspension (E-Z-CAT) to opacify the lumen of the intestine. It is useful in detecting thickening of the bowel wall and in demonstrating the presence of enterocolic and enterovesical fistulas. However, indications for its use in small bowel disorders are limited since the information can often be obtained by standard contrast radiology.

Small bowel enteroscopy

Endoscopy of the small bowel can be performed either by using a long balloon-tipped small bowel enteroscope or by using a 'push' enteroscope. With the former, the balloon is inflated after the endoscope has entered the duodenum and this enables gut peristalsis to carry the tip of the enteroscope to the caecum (Figure 29.5). This is established by radiological screening, and inspection of the bowel is performed as the

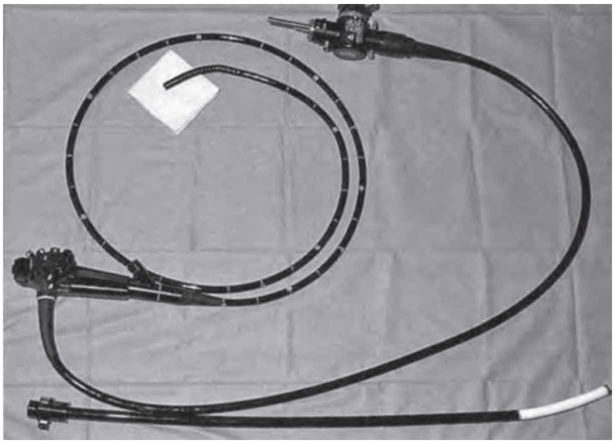


Figure 29.5 Small bowel enteroscope. Note the stiffening overtube at the bottom of the picture.

instrument is slowly withdrawn. This is useful in difficult diagnostic areas such as recurrent obscure bleeding and has shown that many patients with bleeding from non-steroidal anti-inflammatory drugs have duodenal ulcers. The limitations of this type of equipment include the time required to carry out the procedure (in the region of 6 hours) and the inability to biopsy through the enteroscope. The push enteroscope on the other hand can be used to take biopsies, but only the first 60 cm or so of the small bowel beyond the duodenojejunal flexure can be visualized.

A more recent development is double-balloon endoscopy (DBe) and is described together with capsule endoscopy (CE).

Mesenteric angiography

Selective mesenteric angiography can be used to detect angiodysplastic lesions and vascular tumours in the small bowel, which can lead to occult or frank bleeding from the gastrointestinal tract. It is also possible to detect an active bleeding site in the small intestine by this means if blood is being lost at more than 0.5 mL/min. If a bleeding point is identified during mesenteric angiography, it is helpful if the radiologist can place a superselective catheter into the appropriate branch of the SMA to lie as close as possible to the lesion. The catheter is then left *in situ* when the patient goes to the operating theatre. The surgeon can then inject contrast (methylene blue) along the catheter and the segment of small bowel in which the bleeding lesion is present can easily be identified (Figure 29.6). However, mesenteric angiography has been largely superseded by CE and DBe, except in life-threatening bleeding where radiological control by selective embolization of the lesion may be life saving or provide time for resuscitation and urgent surgical intervention.

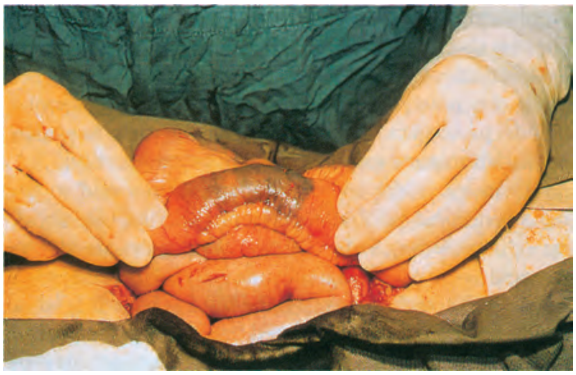


Figure 29.6 Segment of small bowel highlighted by intraoperative injection of methylene blue via superselective mesenteric angiogram catheter.

Capsule endoscopy

CE, introduced in 2000 by Given Imaging is a natural consequence of advances in complementary metal oxide semiconductor (CMOS) sensing technology. CE entails the ingestion of a miniature, pill-sized camera that moves through and visualizes the mucosa of the digestive tract. CE enables inspection of the gastrointestinal tract without discomfort or need for sedation, and thus obviates the risks associated with traditional endoscopy. The first capsule, PillCam SB (Given Imaging, Yoqneam, Israel) was specifically designed for the diagnosis of disorders of the small intestine. The capsule houses a CMOS camera that acquires two images per second during its passive progression through the lumen of the gastrointestinal tract and has a battery life of approximately 8 hours. The visual data obtained by the CMOS camera are sent to a data logger worn by the patient. Following the completion of the CE, the visual data are loaded on to a dedicated PC for viewing (Figure 29.7).

Since then the same company has produced the PillCam ESO for investigation of disorders of the oesophagus and the PillCam Colon for inspection of the colon (see relevant chapters). Within a reasonably short period of time CE has become established in routine clinical practice especially for the investigation of patients with small bowel disorders.

Several large retrospective and prospective clinical studies have confirmed the superior safety and efficacy of CE compared with all other investigations for the diagnosis of obscure gastrointestinal bleeding. It has thus become the investigation of choice in these patients, largely superseding all other investigations. The role of CE in the evaluation of Crohn's disease remains unclear. In contrast, CE has become the key diagnostic test for patients with suspected tumours of the small bowel. There is also good published evidence that CE is useful for the diagnosis of villous atrophy (coeliac disease).

CE has a number of limitations. These include low diagnostic specificity, risk of capsule retention, inability to detect the exact site of the disease, limited battery power, absence of active propulsion and inability to control the orientation of the CMOS camera. The projection is however that, as the technology advances to overcome all the present limitations, CE will ultimately replace conventional endoscopy.

Small bowel CE with PillCam SB

Meticulous bowel preparation is essential for CE to avoid smudging of the camera lens, since the capsule has no irrigation facilities. The small bowel preparation includes ingestion of

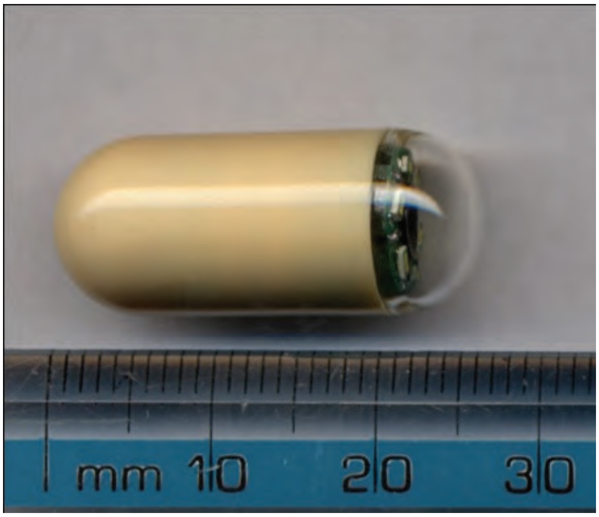


Figure 29.7 PillCam SB: actual capsule which is swallowed by the patient.

1.0 L of clear liquid or polyethylene glycol 12 hours before the examination.

Although the use of prokinetic agents to facilitate propulsion of the capsule and overcome any stoppages at sphincter sites (pylorus), the results of the reported studies on the effect of prokinetics on CE have been contradictory. The largest published series of CE for obscure gastrointestinal bleeding, which included 260 patients, reported a diagnostic yield of CE of 60% and 46% for overt and occult bleeding respectively.

Lesions identified by CE may be treated by DBe (Figure 29.8a) developed by Fujinon (Saitama, Japan). CE and DBe are thus complementary techniques. The complementary role of CE and DBe in the detection of the cause and in the management of obscure gastrointestinal bleeding has been evaluated in a prospective study. This showed a higher sensitivity (90.6% vs 65.6%) and diagnostic yield (71.9% vs 65.9%) for CE than for DBe. However, DBe has the advantage of obtaining biopsy samples and, in suitable cases, successful treatment of lesions. Some practical difficulties have been reported with the use of DBe, including attachment of the balloon to the tip of the endoscope before each examination, and inflation/deflation of the two balloons with every insertion. Additionally, care must be taken during cleaning of the endoscope to avoid blockage of the small air hole at the tip of the device. The single-balloon enteroscopy technique, developed in 2008 by Olympus (Tokyo, Japan), is more manageable than DBe, as it requires the endoscopist to inflate only one balloon (Figure 29.8b). Furthermore, the single-balloon enteroscopy device is easier to clean than the DBe.

Isotope studies

External isotope scintigraphy following the intravenous administration of radiolabelled compounds or isotope-labelled autologous cells can be useful in the investigation of patients with occult gastrointestinal bleeding. It may also be useful in detecting inflamed intestine and in estimating the intestinal transit time.

Intestinal bleeding

Haemorrhage of small intestinal origin may be due to Meckel's diverticulum, polyps, jejunal ulcers, tumours or vascular malformations. These may not be detectable using endoscopy, small barium studies or angiography. Under these circumstances, radionuclide methods may help to solve the problem, although again as with mesenteric angiography they have been largely superseded by CE and DBe when these are available.

Basically there are three isotope methods. The first involves the injection of technetium pertechnetate. This is the method of choice for the detection of bleeding Meckel's diverticulum and carries an accuracy rate of 90%. The technetium is concentrated by ectopic gastric mucosa in the diverticulum which can be identified as a hot spot (Figure 29.9). For the detection of rapidly bleeding sites in the emergency situation ^{99m}Tc sulphur colloid can be used following intravenous bolus injection. It is cleared from the circulation by the macrophage system within 15 minutes, at which point the blood radioactivity declines, but the extravascular radioactivity at the bleeding site increases thereby allowing its detection as a hot spot by external scintigraphy. In practice, this method is rarely used as mesenteric angiography gives a much more accurate anatomical localization and enables therapeutic embolization of the lesion.

In the patient with intermittent gastrointestinal bleeding, however, technetium-labelled autologous red cells may be useful. These are cleared much more slowly from the vascular compartment following their intravenous injection and the method allows repeated examinations over a 24 hour period (Figure 29.10). Its main disadvantage is that it takes time to label the patient's red blood cells and it is therefore unsuitable for the actively bleeding patient.

Estimation of small bowel transit time

In the past, small bowel intestinal transit studies have been performed using radio-opaque solid, non-absorbable chemical markers. These however are generally regarded as being unphysiological and unsuitable for routine clinical practice. Estimation of the small bowel transit is best performed either by external scintigraphy after administration of isotope-labelled meals or by breath tests (see below). Both liquid and solid meals labelled with technetium sulphocolloid or diethylamine penta-acetic acid can be used to estimate simultaneously gastroemptying and small bowel transit time. The detection of caecal radioactivity is used as the end point for the estimation of the small bowel transit time.

Detection of small bowel inflammatory disease

Indium-labelled autologous leucocytes when injected intravenously settle in areas of inflammation or abscess formation. This can be used to detect the extent of active

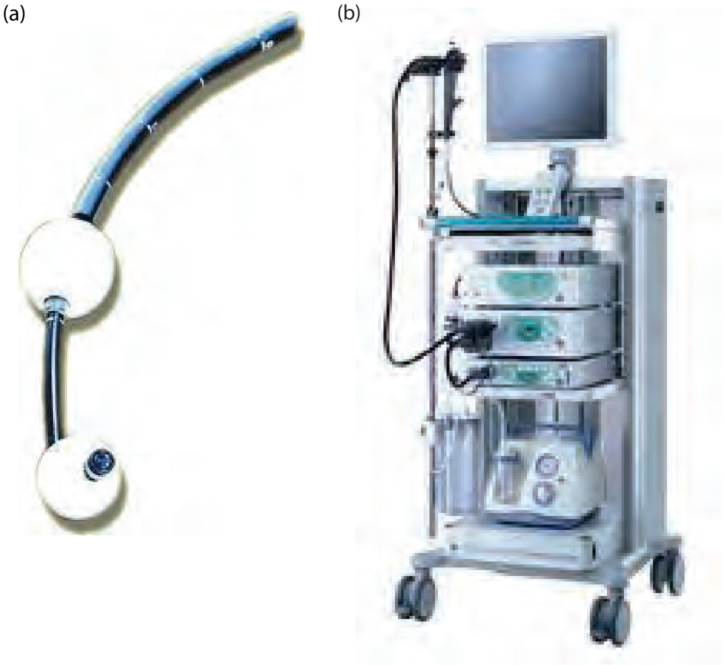


Figure 29.8 (a) Fujinon double-balloon endoscope: enables isolation of a segment between the two balloons which enables treatment of certain small bowel lesions. (b) Olympus single endoscope: also enables treatment of certain small bowel lesions.

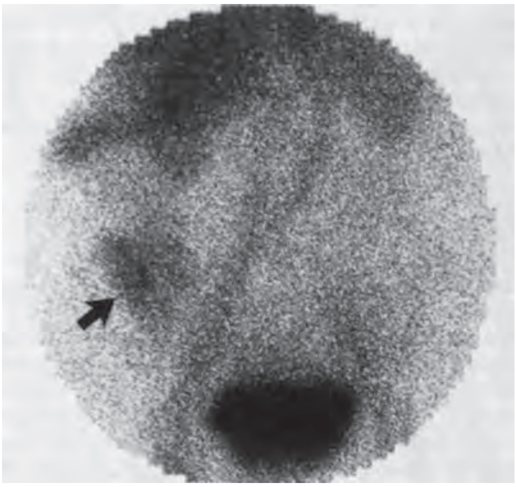


Figure 29.9 Ectopic gastric mucosa in a Meckel's diverticulum outlined by external scintiscanning after the intravenous administration of ^{99m}Tc.

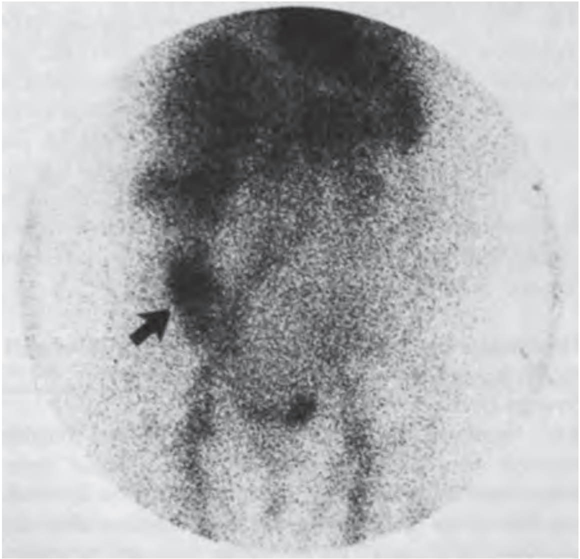


Figure 29.10 Occult bleeding lesion in the ascending colon located by the ^{99m}Tc-labelled autologous red cell technique. Selective arteriography was unhelpful in this elderly patient. The lesion proved to be an angiodysplasia.

inflammatory bowel disease and assess its severity. However, there can be great difficulty with interpretation because of the increased background radioactivity produced by the uptake of the radiolabelled leucocytes by the bone marrow, liver and spleen. Another approach is to use labelled sucralfate. Sucralfate is an aluminium salt of polysulphated sucrose which is used in the treatment of peptic ulceration. It binds selectively to areas of mucosal ulceration within the gastrointestinal tract and labelled sucralfate has proved to be a useful technique in the detection of inflammatory bowel disease in both the large and the small intestine. In view of its lower radiation dosage than barium studies, it can be used as a screening test and in the serial assessment of disease activity. The labelled suspension of sucralfate is administered by mouth, and serial isotope scans of the abdomen are carried out at 2, 6, 20 and 24 hours.

Investigations for malabsorption

Malabsorption implies the inability to absorb sufficient nutrient owing to disease of the gastrointestinal tract. It is most commonly due to dysfunction of the small bowel or the pancreas. The most common clinical feature is steatorrhoea from malabsorption of fat. Symptoms attributable to failure to absorb carbohydrate and protein are not so common and may consist of abdominal discomfort and bloating. Digested carbohydrate is fermented to lactic acid in the large bowel, which can lead to diarrhoea. The investigation of a patient with suspected malabsorption should start with the estimation of faecal fat. Once steatorrhoea is confirmed further tests are required to establish the nature of the underlying pathology. These tests help to differentiate whether the malabsorption is of small bowel origin (e.g. small intestinal disease or bacterial overgrowth) or the result of pancreatic exocrine insufficiency.

Estimation of faecal fat

The quantitative estimation of faecal fat remains the most sensitive and reliable test of disorders of digestion and absorption. The faecal fat output per day is estimated on a 3–5 day collection on a standard diet containing 80–100 g of fat. The normal is less than 6.0 g per day (80 mmol of triglyceride). Other tests such as the ^{14}C-triolein breath test and oxalate-loading test are less reliable but are used in some centres as screening tests for steatorrhoea.

Duodenal mucosal biopsy

Biopsy taken at upper-gastrointestinal endoscopy of the distal duodenum can be used to look for abnormalities of villous architecture, e.g. subtotal villous atrophy in coeliac disease (Figure 29.11). In addition, abnormal mucosal pathogens may be detected as in Whipple disease in addition to parasites, e.g. *Giardia lamblia* (Figure 29.12). Endoscopic duodenal biopsy has largely replaced the use of the suction Crosby capsule.

The Schilling test

The absorption of vitamin B_{12} by the terminal ileum requires the presence of intrinsic factor and to a lesser extent the R protein in the gastric juice. In the Schilling test radiolabelled vitamin B_{12} (1.0 μg) is administered orally immediately after a large parenteral injection of the unlabelled vitamin (1000 μg) to ensure saturation of the body stores. Under these conditions, normal subjects will excrete 10% or more of the radiolabelled vitamin in their urine. If abnormally low excretion is found in a patient, the test is repeated but the labelled vitamin B_{12} is given together with intrinsic factor. In the presence of ileal disease, the abnormally low excretion of the labelled vitamin in the urine is not altered by the addition of the intrinsic factor. However, in patients with pernicious anaemia or after total gastrectomy the administration of intrinsic factor restores the urinary excretion of the labelled vitamin to normal. Both stages of the test are invalidated by dehydration and renal disease. Bacterial overgrowth may cause malabsorption of the vitamin and an abnormal Schilling test, but this will revert to normal after a course of antibiotic therapy.

SeHCAT bile absorption test

This test estimates the ability of the terminal ileum to absorb bile acids. The synthetic selenium bile acid known as SeHCAT is used. A dose of this ^{75}Se-labelled compound is administered orally or intravenously and a gamma-counter is used to estimate the bile acid absorption. There is a very good correlation between the results of the SeHCAT test and faecal excretion of bile acids.

^{14}C-glycocholate and the ^{14}C-D-xylose breath tests

The ^{14}C-glycocholate breath test is used to detect bacterial overgrowth in the small intestine. The glycine moiety of the conjugated bile salt is labelled with ^{14}C and ingested to mix with the endogenous bile salts in the intestine. Normally the bile salts are largely reabsorbed intact in the terminal ileum to enter the enterohepatic circulation and only a small amount reaches the colon, where it is deconjugated by the colonic bacteria and the glycine metabolized to yield $^{14}CO_2$. This is absorbed and eliminated in the expired air. However, in the presence of bacterial overgrowth most of the ingested labelled salt is deconjugated by the small bowel bacteria and excess $^{14}CO_2$ is produced and eliminated in the expired air (Figure 29.13). False-positive results are obtained with this test in the presence of ileal disease. The ^{14}C-D-xylose test is more reliable in this respect. D-Xylose is a pentose which is normally absorbed intact by the same transport mechanism as the hexoses.

Breath tests for carbohydrate malabsorption

The analysis of breath $^{14}CO_2$ following the ingestion of ^{14}C-lactose is a convenient test for lactose intolerance resulting from a deficiency of the brush border enzyme lactase. This test is easy to perform, is as accurate as the lactose tolerance test and agrees reasonably well with mucosal disaccharidase activity. Hydrogen is produced when carbohydrate is fermented by some bacteria. When there is lactose malabsorption the sugar reaches the colon, where it is fermented with the production of hydrogen. A proportion of this gas is absorbed and excreted in the expired air. Using mass spectrometry it is possible to measure very low concentrations of hydrogen in a sample of end-expiratory air which has a similar composition to that of alveolar air. There is evidence that the breath hydrogen concentration is more accurate than $^{14}CO_2$ excretion for the diagnosis of lactase deficiency. Both tests give false-positive results in patients with bacterial overgrowth (Figure 29.14).

Hydrogen breath tests for measurements of small bowel transit time and bacterial overgrowth

The hydrogen breath test is a useful and reliable method for determining small bowel transit time. Repeated measurements of the hydrogen in the end-expiratory air are taken every few minutes after the ingestion of a meal. The latter may be liquid in nature (drink of the non-absorbable sugar lactulose) or solid, usually mashed potatoes and baked beans, which contain non-absorbable oligosaccharides. When the meal

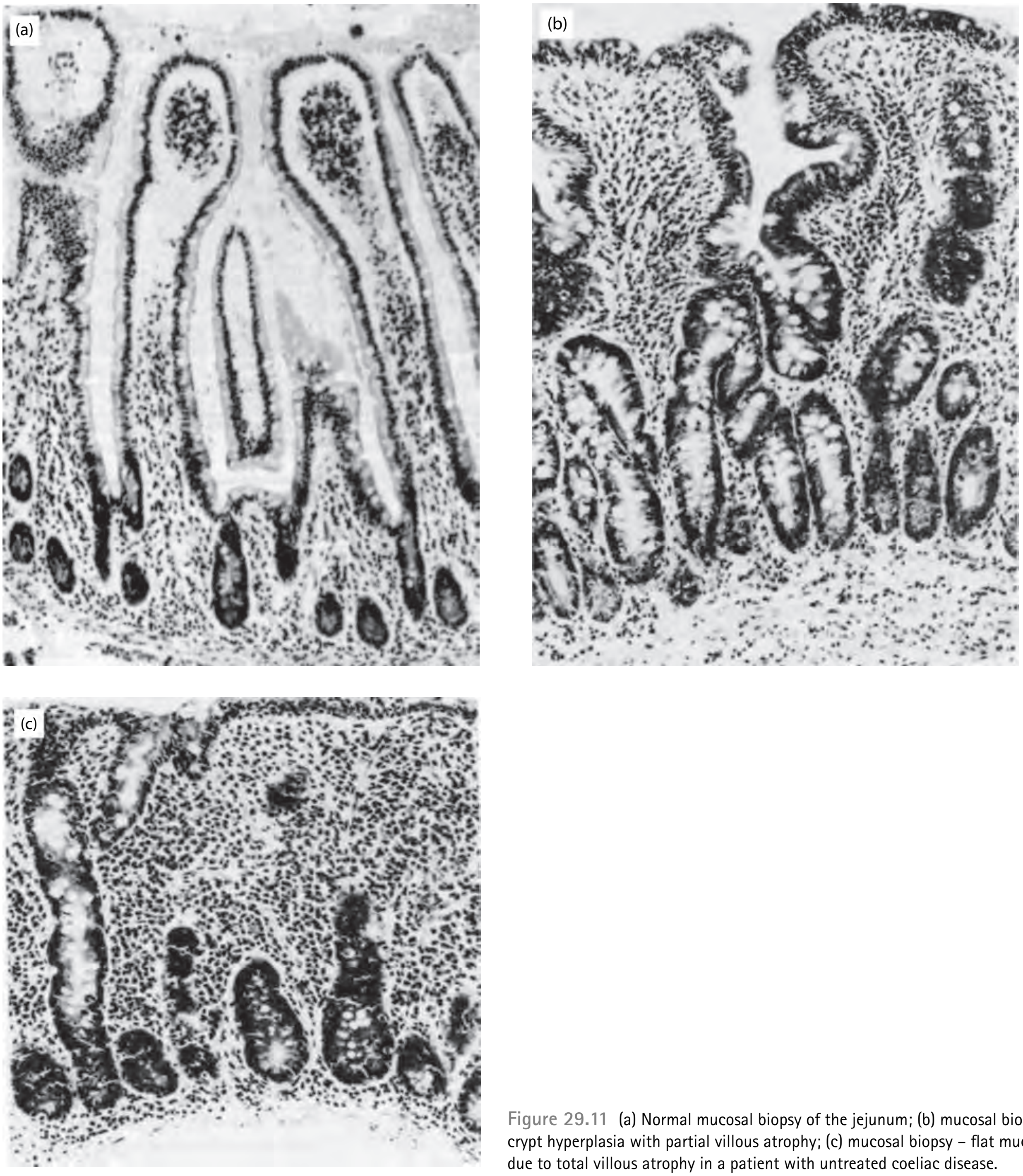


Figure 29.11 (a) Normal mucosal biopsy of the jejunum; (b) mucosal biopsy – crypt hyperplasia with partial villous atrophy; (c) mucosal biopsy – flat mucosa due to total villous atrophy in a patient with untreated coeliac disease.

reaches the caecum the resulting bacterial fermentation induces a sustained rise in the breath hydrogen concentration (Figure 29.15). Strictly speaking, the test measures the oral–caecal transit time, which includes the gastric emptying time. However, if the meal is radiolabelled with technetium both gastric emptying and small bowel transit times can be calculated from a single investigation.

In patients with bacterial overgrowth the fasting hydrogen concentration in the expired breath is elevated. In addition, there is an early rise in the hydrogen in the expired air following the administration of the lactulose.

Tests of intestinal permeability

Various disorders including Crohn's disease and mucosal enteropathies are associated with abnormal permeability of the intestinal mucosa to macromolecules that are not absorbed by the intact normal mucosa. The methods used to test permeability are based on the absorption of substances

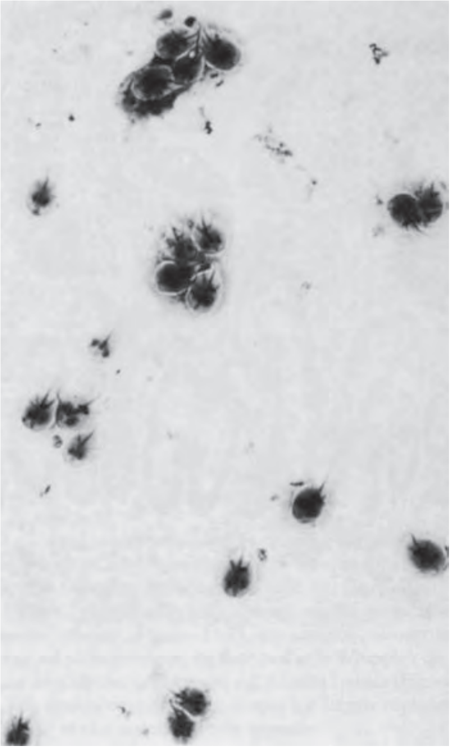


Figure 29.12 *Giardia lamblia* obtained from a jejunal mucosal biopsy.

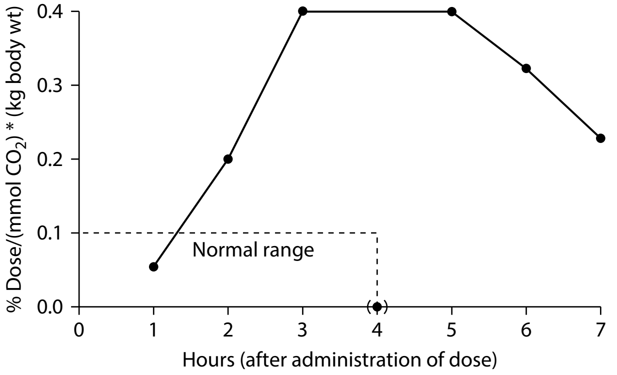


Figure 29.13 ^{14}C-glycocholate breath test in a female patient with weight loss, hypoproteinaemia and moderate steatorrhoea. The investigation is clearly abnormal with an increased amount of labelled CO_2 being detected in the expired air. Normally less than 0.1% of the administered dose is recovered in the expired breath. The patient had an ileotransverse anastomosis for Crohn's disease. The bacterial overgrowth subsided when the excluded diseased bowel was resected.

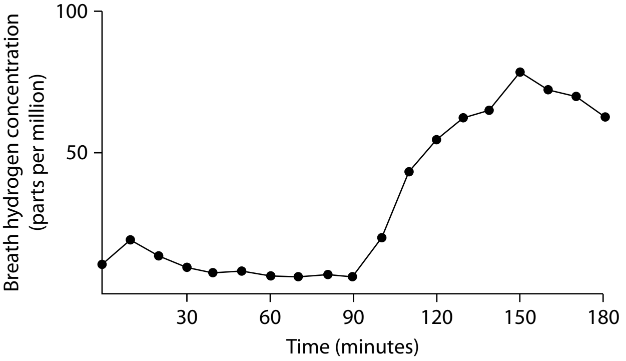


Figure 29.14 Breath hydrogen excretion in a patient with small bowel bacterial overgrowth. Breath hydrogen excretion in a normal subject following 10 g of lactulose. Determination of the oral–caecal transit time by the hydrogen breath test. Following the administration of lactulose solution, serial H_2 estimations are performed on samples of end-expiratory air. A sustained rise in the H_2 in the expired air indicates that the head of the meal has reached the caecum where bacterial fermentation of the non-absorbable carbohydrate occurs.

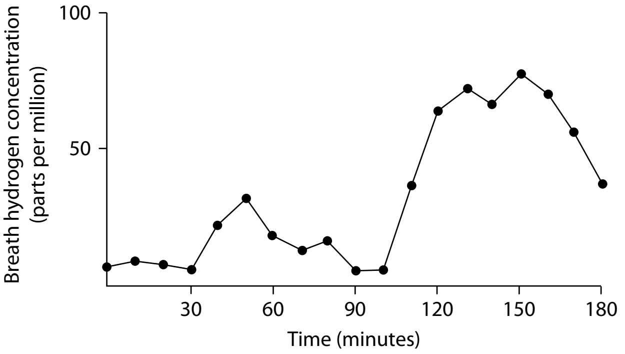


Figure 29.15 Breath hydrogen excretion in a patient with small bowel bacterial overgrowth. Breath hydrogen excretion in a normal subject following 10 g of lactulose. H_2 lactulose breath test in a patient with bacterial overgrowth. The fasting H_2 in the expired air is high and there is an early rise in the breath H_2 after the ingestion of the lactulose solution.

that are normally excluded and if absorbed (because the gut permeability is pathologically enhanced) are not biodegraded and can thus be detected in the blood or urine. Clinically the pathological absorption of these substances is detected by analyses of urine samples for the specific agent used in the test. Examples of test substances include polyethylene glycol 400 and combinations of small and large compounds with differential absorption, e.g. mannitol and cellobiose. When differential studies using two substances are used, the abnormal permeability is demonstrated by changes in the ratio of the urinary concentrations of high- to low-molecular-weight test compounds. The oral administration of ^{51}Cr-labelled EDTA can be used as a quick non-invasive test of abnormal mucosal permeability.

Small bowel neoplasia

Small bowel tumours are rare and collectively account for less than 10% of all gastrointestinal neoplasms. The majority (70%) of small bowel neoplasms are benign and include epithelial tumours (adenomas), lipomas, haemangiomas, gastrointestinal stromal tumours (GISTs) and neurogenic tumours. A number of gastrointestinal disorders are associated with an increased risk of invasive malignancy. These include Crohn's disease, coeliac disease, dermatitis herpetiformis and radiation enteritis. Additionally, invasive malignant transformation occurs with certain polyposis syndromes (see below).

Benign small bowel tumours

Many benign small bowel tumours remain asymptomatic throughout life. Others present either with gastrointestinal bleeding or with intestinal obstruction, most commonly due to intussusception. This should be distinguished from idiopathic intussusception, which commonly occurs in children under the age of 2 years. Tumour-induced intussusception occurs later than the idiopathic variety and does not usually reduce by hydrostatic treatment. Chronic blood loss from a benign small bowel tumour causing iron deficiency anaemia is a fairly common presentation of small bowel tumours and the more vascular tumours may give rise to acute severe gastrointestinal bleeding. If available, small bowel CE has replaced most other investigative tests in patients with this presentation. Otherwise, the diagnosis may be made by small bowel follow through or small bowel enema or, in the case of vascular tumours, by mesenteric angiography. In patients presenting with intestinal obstruction, the diagnosis is usually established at emergency laparotomy. Benign small bowel tumours may be single or as multiple.

Surgical excision of small bowel tumours is the standard treatment. Exploratory laparotomy with total excision of the lesion provides the safest treatment and material for establishing a definite diagnosis, especially for lesions which may be malignant. Tumours discovered incidentally during laparotomy for surgical treatment of other symptomatic disease should be removed and subjected to detailed histological examination. Certain lesions discovered because of occult bleeding may be treated endoscopically with the double or single balloon operating endoscope. Acute life-threatening bleeding may be managed by angiographic embolization with open surgery in reserve should embolization fail. All patients who develop intestinal obstruction are managed with emergency laparotomy. Segmental resection and enterotomy/polypectomy can be used depending on the extent and pathology of the disease. If the pathology cannot be established at the time of resection, full segmental resection with adequate margins is recommended.

Hyperplastic polyps

These are benign mucosal single or multiple lesions occurring in the duodenum, jejunum and proximal ileum. The duodenal lesions are frequently discovered by routine upper gastrointestinal endoscopy. Hyperplastic polyps are generally asymptomatic with no malignant potential. They may be removed endoscopically with biopsy forceps or with the single or double balloon endoscope if situated further distally.

Adenomas

There are of three types: (1) adenomatous polyps, (2) Brunner gland adenomas and (3) villous adenomas. All may be single or multiple, and sessile and pedunculated. Histologically, they appear as intraluminal extensions of the mucous membrane and submucosa with multiple acini supported on a central fibrovascular core. Varying degrees of differentiation are encountered with the various types. Adenomas may cause obstruction, bleeding, intussusception and, occasionally, malignant degeneration, particularly with larger villous lesions. Brunner gland adenomas develop most often along the posterior wall of the duodenum at the junction of the first and second parts and may be single, multiple or diffuse. Histologically, they exhibit benign proliferation of the Brunner glands with scattered ductal and stromal elements.

Villous adenomas are rare and most frequently found in the duodenum. They may give rise to bleeding or obstruction, and as with their counterparts in the colon and stomach, they may be associated with malignant degeneration. Villous adenomas larger than 4.0 cm are at particular risk of developing or harbouring malignancy.

Mesenchymal tumours

These include GISTs, true gastrointestinal smooth muscle tumours (GISMTs) and neurogenic tumours. GISTs are identified on the basis of their immunohistochemical staining characteristics. All are positive for c-Kit (CD117) and CD34 and although some GISTs show positive staining for muscle actin, all are negative for desmin and S100 protein. With all mesenchymal tumours, the distinction between the benign and malignant forms is often difficult. In general however malignant tumours are larger, more often ulcerated and exhibit marked cellularity and necrosis. The tumour can be confidently labelled as benign only if the patient is disease free for at least 3 years after surgical excision. Smooth muscle tumours may occur anywhere in the small intestine but are more commonly found in the jejunum and ileum.

Gastrointestinal stromal tumours or gut stromal tumours

GISTs are the most common mesenchymal neoplasms of the gastrointestinal tract and can also originate in the mesentery and omentum. Overall, however, GISTs are rare, with a lower prevalence than adenocarcinomas and lymphomas of the gastrointestinal tract. They may form intraluminal or extraluminal or transmural (dumbbell-shaped) tumours.

The term GIST was coined by Mazur and Clark in 1983, as these tumours do not exhibit the ultrastructural and immunohistochemical features characteristic of smooth muscle. Instead, they are thought to originate from pluripotential mesenchymal stem cells which normally differentiate into the interstitial cell of Cajal (gastrointestinal pacemaker cells, responsible for initiating and co-ordinating gastrointestinal motility), hence the alternative name suggested by Kindblom *et al.* of ‘GI pacemaker cell tumours’. GISTs can occur anywhere in the gastrointestinal tract. They originate as submucosal tumours which usually grow endoluminally in the affected viscus, although instances of GISTs with predominant extraluminal growth are well documented. At the time of diagnosis GISTs range in size from 1 to 40 cm and above. The stomach is the most common site of origin (50–70%) followed by the small intestine (20–30%). Less frequent sites include the colon and rectum (5–15%), oesophagus (<5%) and very rarely duodenum, omentum and intestinal mesentery. GISTs have a tendency to surface ulceration and deep necrosis, giving rise to severe acute bleeding (the most common presentation of GIST), although minor occult bleeding from surface erosion may occur and give rise to iron-deficiency anaemia. Other complications of GIST include bowel obstruction, intussusception and tumour perforation. Differentiation between benign and malignant GISTs can be difficult even on histological examination.

Molecular biology studies have led to the reclassification of GISTs as distinct tumours separate from smooth muscle tumours, and provided information on their molecular pathogenesis, which, in turn, has led to newer more effective medical therapy based on inhibition of tyrosine kinases as these tumours are largely unresponsive to conventional chemotherapy.

Aetiology and pathology

Mutations in exon 11 of the c-*kit* proto-oncogene are associated with most GISTs. These mutations cause overexpression and autophosphorylation of c-Kit, inducing a cascade of intracellular signalling that stimulates cell proliferation. The *c-kit* proto-oncogene is located on chromosome arm 4q11–12. It encodes KIT, which is a transmembrane tyrosine kinase. Stem cell factor acts as the ligand for KIT. Under normal circumstances, KIT activation occurs when stem cell factor binds to the extracellular domain of c-Kit resulting in homodimerization of the normally inactive c-Kit monomers. The resulting autophosphorylation of intracellular tyrosine residues exposes the binding sites for the intracellular signal transduction molecules with activation of a signalling cascade that involves phosphorylation of several downstream target proteins, including MAP kinase and RAS. Following signal transduction to the nucleus, marked mitotic activity and protein transcription ensue.

Although some families with hereditary GISTs have been described, the majority of cases are sporadic. One of these rare familial types is characterized by multiple GISTs with or without the presence of dermal and mucous membrane hyperpigmentation, multiple naevi and urticaria pigmentosa, mast cell dysfunction and diffuse hyperplasia of gastrointestinal spindle cells. GISTs occur with a higher than expected frequency in patients with type 1 neurofibromatosis and in patients with the rare *Carney triad* (epithelioid gastric stromal tumours, pulmonary chondromas, and extra-adrenal paragangliomas), which usually affects young females.

On histopathological examination, three types of GISTs are recognized: spindle cell (70%), epithelioid cell (20%) and mixed types (10%), with the spindle cell type being particularly difficult to differentiate from smooth muscle tumour although these lack c-Kit immunoreactivity. The Fletcher *et al.* histological classification of GISTs is useful as it differentiates between very low-, intermediate- and high-risk tumours with corresponding different prognosis: 5 year survival rates of 34% for high-risk tumours and 63% for the low-risk groups (Table 29.2).

Table 29.2 Fletcher *et al.* histological classification of gastrointestinal stromal tumours

	Size (cm)	**Mitotic count/50 HPFs**
Very low risk	<2	<5
Low risk	2–5	<5
Intermediate risk	<5	6–10
	5–10	<5
High risk	>5	>5
	>10	Any
	Any	>10

HPF, high-power fields

The pathology report should also include predominant cell type, extent of nuclear pleomorphism, extent of tumour, distance from resection margins and lymph node deposits. The most common sites of metastatic disease from malignant GISTs are the liver (71%) and peritoneum (53%).

Clinical features

The reported incidence of GISTs is one to two per 100 000 per annum. The age incidence of GISTs is 55–65 years, although they may occur in young adults. Most GISTs are diagnosed when they are less than 4.0 cm in diameter, frequently as an incidental finding during endoscopic or radiological investigations or during elective or emergency surgery for gastrointestinal haemorrhage or intestinal obstruction. Lesions greater than 5.0 cm in diameter are more likely to cause symptoms which include:

- non-specific abdominal pain or discomfort
- early satiety or a sensation of abdominal fullness
- rarely, a palpable abdominal mass
- obstruction: from intraluminal growth of an endoluminal tumour or intussusception induced by a small submucosal tumour or from extrinsic luminal compression by an exophytic lesion. The obstructive symptoms can be site-specific: dysphagia (oesophageal GIST), acute small bowel obstruction (jejunoileal GIST), constipation/chronic obstruction (colorectal GIST), obstructive jaundice (duodenal GIST)
- gastrointestinal haemorrhage: from pressure necrosis and ulceration of the overlying mucosa
- perforation with acute peritonitis: usually from tumours of the stomach and small intestine. The perforation may be silent and localized with late presentation as an intra-abdominal abscess
- symptomatic anaemia in patients with recurrent occult blood loss: malaise, fatigue or exertional dyspnoea
- in some patients the tumour is discovered during investigation for an unrelated disorder or as an unexpected finding during emergency surgery for acute bleeding or perforated viscus.

In the absence of complications or large symptomatic tumours (palpable abdominal mass), the physical examination is usually negative as there are no physical findings to specifically suggest the presence of a GIST. Some patients present with physical findings associated with gastrointestinal blood loss, bowel obstruction, bowel perforation and abscess formation.

Surgical treatment

It is important to stress that the optimum management of GISTs requires close participation between the surgeon and medical oncologist. Despite the recent advances in tyrosine kinase inhibitor therapy, surgery remains the definitive therapy for patients with GISTs as radical and complete surgical excision offers the only chance for cure. Surgical treatment may also be indicated in symptomatic patients with locally advanced or metastatic disease, since debulking of large tumours improves the response obtained with adjuvant imatinib mesylate therapy. In patients presenting with acute complications (major gastrointestinal bleeding, intestinal obstruction, peritonitis), resuscitation and stabilization is followed by emergency surgical intervention. Hepatic artery embolization and chemoembolization may also be considered in patients with imatinib-resistant tumours with progressive liver metastases.

There is growing evidence that the laparoscopic surgical approach should be considered for the resection of gastric GISTs. In a seminal study by Novitsky *et al.*, 50 consecutive patients undergoing laparoscopic resection of gastric GISTs were identified in a prospectively collected database. Outcome measures included patient demographics and outcomes, operative findings, morbidity and histopathological characteristics of the tumour. Patient and tumour characteristics were analysed to identify risk factors for tumour recurrence. The patients underwent 47 local and three segmental laparoscopic gastric resections. Mean tumour size was 4.4 cm with the majority of the lesions located in the proximal stomach. The mean operative time was 135 minutes, the average blood loss was 85 mL and the mean length of hospitalization was 3.8 days. There were no major perioperative complications and no mortality. All the resections had negative resection margins. At a mean follow-up of 36 months, 46 (92%) patients were disease free, one patient was alive with disease, one patient with metastases died of a cardiac event and two (4%) patients died of metastatic disease. No local or port site recurrences were identified. Patient age, tumour size, mitotic index, tumour ulceration and necrosis were statistically associated with tumour recurrence. As reported in other series, the presence of 10 or more mitotic figures per 50 high-power fields (HPFs) was an independent predictor of disease progression.

Medical treatment

Conventional systemic chemotherapy is not effective in GISTs and has largely been replaced by tyrosine kinase inhibitors. Intraperitoneal chemotherapy with cisplatin and doxorubicin or mitoxantrone is nowadays used as salvage therapy in patients resistant to tyrosine kinase inhibitor therapy. GISTs are radioresistant, and for this reason radiotherapy is not indicated. The only effective, specific, cytotoxic is imatinib mesylate (Gleevac). This is a selective tyrosine kinase inhibitor with action against mutant c-Kit. Druker *et al.* first reported promising response rates following therapy with imatinib in patients with chronic myeloid and acute lymphoblastic leukaemia. This was followed by several phase II clinical trials with imatinib including a large series of patients with GISTs reported by Blanke *et al.* involving 145 patients. Although in this study no complete responses were observed,' partial clinical response was obtained in 59% of patients on treatment with imatinib in doses of 400–600 mg per day. A subsequent series reported by van Oosterom *et al.* on 35 patients with GISTs treated with imatinib in doses of 400–1000 mg per day obtained partial response rates in 54%, and disease stability with no progression in a further 37%. The most common side effects reported by patients are periorbital and peripheral oedema, fatigue, skin rash and nausea and vomiting. At 10 months of follow-up, 82% of patients maintained the partial clinical response or showed no evidence of disease progression. However, subsequent studies with longer follow-up have shown that, despite the initial efficacy, many patients develop drug resistance to imatinib mesylate. The most effective dose of imatinib mesylate remains unknown pending results of randomized prospective clinical trials. The current recommended dose in adult patients is 400 mg orally q.i.d. with food.

Sunitinib (Sutent) is a multikinase inhibitor that targets several tyrosine kinases implicated in tumour growth, angiogenesis and metastatic progression. It inhibits platelet-derived growth factor receptors (PDGFR-α and -β), vascular endothelial growth factor receptors (VEGFR1, VEGFR2 and VEGFR3), stem cell factor receptor (KIT), Fms-like tyrosine kinase-3 (FLT3) and colony-stimulating factor receptor type 1. Sunitinib is indicated in patients with GISTs in whom the disease has progressed or who are resistant or unable to tolerate treatment with imatinib. The reported clinical studies show that it delays median time to tumour progression. The adult dose is 50 mg orally q.i.d. for 4 weeks with a rest for 2 weeks before the cycle is repeated. Other tyrosine kinase inhibitors known to be effective in leukaemias are currently being tested for their efficacy in patients with advanced GISTs unresponsive to Gleevec and Sutent.

Prognosis

The prognosis depends on extent of the disease (localized primary vs metastatic/recurrent disease) and, in particular, on the histopathological features of the tumour. Hence the overall 5-year survival rate varies widely from 25% to 65%. Larger GISTs are associated with complications, e.g. gastrointestinal haemorrhage, gastrointestinal obstruction and bowel perforation. Individual tumours are classified into high- and low-risk categories based on size and mitotic activity. In the published literature benign GISTs are generally two to three times more common than the malignant counterparts. Histological features suggestive of malignancy include large tumour cell size, lack of cellular differentiation and more than 10 mitotic figures per 50 HPFs (see Fletcher *et al.* classification).

Gastrointestinal smooth muscle tumours

Although smooth muscle tumours may affect other organs such as the uterus and urinary bladder, etc., only GISMTs are considered here. In the past prior to advances in molecular oncology, many GIST tumours were misdiagnosed as GISMTs. It is now established that these tumours are much rarer than

GISTs, with an estimated relative frequency of one GISMT for 36 GISTs. By definition on histology, the component tumour cells resemble smooth muscle with visible, often fibrillar and eosinophilic cytoplasm, usually in long bundles or fascicles. The nucleus is long and oval, often with a cytoplasmic vacuole and indistinct nucleoli. In addition, there is some deposition of collagen between smooth muscle cells, which may resemble myofibroblasts (less cytoplasm, smaller nuclei with dense chromatin). Histochemistry with trichrome stain colours the smooth muscle cytoplasm red and the collagen blue or green. All GISMTs are negative on testing for c-Kit immunoreactivity.

GISMTs are rare constituting only 1% of gastrointestinal tumours, but they occur with an increased incidence in immunocompromised patients (AIDS, transplanted patients and congenital immune deficiency disorders) compared with the general population, as all these patients are at risk of immunosuppression-associated malignancies which include smooth muscle tumours. In these immunodeficient patients, smooth muscle tumours are strongly associated with the immune dysregulation caused by the Epstein–Barr virus (EBV).

Most GISMTs are benign (leiomyomas, ≤5 mitoses per 50 HPFs) but some (14–14%) are malignant (leiomyosarcomas, ≥35 mitoses per 50 HPFs). GISMTs can occur in the oesophagus and rectum, but the commonest site is the stomach (body and antrum) and they are less common in the small intestine. Some are asymptomatic and discovered as incidental findings during investigation of an unrelated disorder. The clinical presentation in symptomatic patients depends on tumour location, but vague abdominal pain and gastrointestinal bleeding are the usual presenting symptoms. Complete surgical excision is the treatment of choice and is always curative for leiomyomas. Leiomyosarcomas do not respond to radiotherapy or chemotherapy. The only effective treatment is therefore surgical excision with a wide healthy margin together with the regional lymph nodes and associated mesentery. The prognosis after resection of leiomyosarcoma is however poor with frequent local recurrence and metastases mainly to the liver leading to death in at least 50%

Lipomas

Small bowel lipomas are benign submucosal tumours of mesenchymal origin. These tumours are often located in the ileum and may frequently develop as pedunculated or sessile submucosal lesions. They may grow undetected to a size sufficient to produce symptoms of colicky abdominal pain and intermittent bowel obstruction. Intussusception may be produced by pedunculated lipomas. Histological features include collections of mature adipose tissue and fibrous tissue strands. Evidence of surface ulceration, central necrosis and haemorrhage may be present. Collections of adipose tissue may be found near the ileocaecal valve. These deposits must be differentiated from genuine small bowel lipomas.

Intra-abdominal neurogenic tumours

Although they involve other intra-abdominal organs and spaces much more often than the small intestine, neurogenic tumours are conveniently discussed here. They are broadly classified into:

- nerve sheath tumours: neurilemmomas, neurofibromas, neurofibromatosis
- tumours arising from ganglion cells: ganglioneuromas, ganglioneuroblastomas, neuroblastomas
- tumours arising from the paraganglionic system: phaeochromocytomas, paragangliomas.

By far neurofibromas are the commonest and may be solitary or multiple and associated with systemic neurofibromatosis and café-au-lait skin patches (*von Recklinghausen disease*). Intra-abdominal neurogenic tumours are most commonly located in the retroperitoneum, in the paraspinal regions and in the adrenal glands, although rarely they can occur in other sites, e.g. urinary bladder, bowel wall, gallbladder and abdominal wall. All intra-abdominal neurogenic tumours with the exception of neuroblastomas and ganglioneuroblastomas present in adult patients. On CT, neurogenic tumours have a distinctive appearance as smooth or lobulated masses, often with areas of calcification. In the absence of distant metastases, the differentiation between benign and malignant lesions is often difficult. In suspect lesions, special imaging modalities such as MRI and positron emission tomography scanning may be necessary for characterization, extent of tumour and related staging. The mainstay of treatment of intra-abdominal neurogenic tumours is surgical excision irrespective of type, provided there is no evidence of distant metastases on the preoperative work-up. The resection can be carried out laparoscopically except for large or frankly malignant lesions.

Nerve sheath tumours

These include neurilemmoma, neurofibroma, neurofibromatosis, and neurogenic sarcoma (malignant schwannoma). More than 90% of nerve sheath tumours are benign and usually present in young and middle-aged adults with a female–male incidence of 2:1. Nerve sheath tumours are multiple in 10% of cases.

Neurilemmomas (schwannomas) arise from the nerve sheaths of peripheral nerves as encapsulated lesions with the nerve fibres stretched over and around the tumour. Although they may occur anywhere, they have a predilection for the head and neck region and the flexor surfaces of the limbs. However, they may develop as internal lesions of the retroperitoneum, where they account for 6% of primary tumours of this space and in the posterior mediastinum.

Retroperitoneal tumours are usually located in the paravertebral region adjacent to the kidney or in the presacral space. At presentation, neurilemmomas are usually less than 5.0 cm in diameter, but much larger lesions may present in the retroperitoneal space. These tumours often have areas of central necrosis due to inadequate blood supply with cyst formation, calcification, haemorrhage and hyalinization. Histologically, neurilemmomas consist of two different components: Antoni types A and B. Antoni type A areas are highly cellular and composed of spindle cells, whereas in Antoni type B areas the tumour cells are separated by abundant oedematous fluid, often forming cystic spaces. They present in young to middle-aged adults, with females being affected twice as often as males. The vast majority of these tumours are asymptomatic and hence discovered incidentally. The CT appearance is that of a well-demarcated round or oval mass often with cystic degeneration and calcification (punctate, mottled, or curvilinear along the periphery of the lesion).

Neurofibromas and neurofibromatosis

About 90% of neurofibromas are solitary. The rest occur as part of von Recklinghausen disease. On histology, they consist of proliferation of nerve sheath cells interspersed with thick collagen bundles and may exhibit variable myxoid degeneration. Unlike neurilemmomas, neurofibromas are not encapsulated and do not exhibit a clear partition into Antoni A and B areas. Additionally, neurofibromas are solid tumours and seldom exhibit gross cystic change, although they may undergo myxoid degeneration with formation of multiple small cysts. Neurofibromas may undergo malignant degeneration, particularly in cases of neurofibromatosis, whereas neurilemmomas seldom, if ever, undergo malignant degeneration.

Neurofibromas occur more frequently in men than in women usually in the second to fourth decades of life. On CT scanning they have a homogeneous, smooth, round appearance with a well-defined outline.

The term neurofibromatosis comprises a group of heterogeneous disorders that include two distinct types:

- NF-1 neurofibromatosis
- NF-2 neurofibromatosis: bilateral acoustic schwannomas and multiple meningiomas (central neurofibromatosis).

NF-1 is an inherited disorder, which predominantly affects the skin and peripheral nervous system and is characterized by the presence of café-au-lait spots and neurofibromatous nodules along the peripheral nerves (*von Recklinghausen disease*) (Figure 29.16).

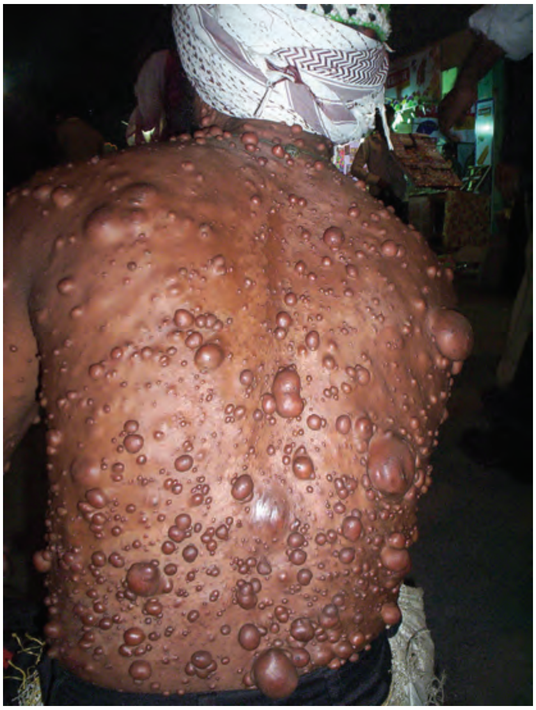


Figure 29.16 NF-1 neurofibromatosis (von Recklinghausen disease). (Courtesy of Professor Omar Shah Srinagar, Kashmir India.)

These extracranial tumours may be solid or plexiform neurofibromas. Approximately 30% of patients with solitary neurofibroma have NF-1, and virtually all patients with multiple neurofibromas or plexiform neurofibromas have NF-1. The abdominopelvic manifestations of NF-1 may involve numerous intra-abdominal ganglia and plexuses, which innervate the abdominal and pelvic organs. These tumours tend to arise in the retroperitoneal, mesenteric and paraspinal regions. Plexiform neurofibromas within the pelvis may form large, infiltrating masses in the presacral or gluteal regions. In contrast, the hallmarks of NF-2 are bilateral vestibular schwannomas and multiple meningiomas. NF-2 rarely demonstrates CNS involvement.

Malignant nerve sheath tumours

Malignant nerve sheath tumours may be encountered in patients with or without neurofibromatosis, but malignant lesions arising in the absence of neurofibromatosis are commoner. Malignant lesions have an equal sex incidence, with the majority arising in adults aged 20–50 years. Differentiation from benign tumours by preoperative imaging tests is unreliable, although irregular borders and lack of homogeneity of the mass are suggestive but inconclusive. More reliable signs of malignancy include large size, bilateral lesions and attenuation by parapsoas masses and their progressive enlargement on serial CT scans. Pain is also common in malignant tumours.

Tumours of ganglion cell origin

These include benign ganglioneuroma, malignant neuroblastoma and ganglioneuroblastoma, the latter exhibiting an intermediate biological aggressive behaviour. The adrenal gland is the most common primary site of these tumours. Neuroblastoma and ganglioneuroblastoma usually occur in infants and children, whereas ganglioneuroma usually affects adolescents and young adults.

Ganglioneuromas

Ganglioneuromas are rare benign tumours arising from sympathetic ganglia. They are composed of mature Schwann cells, ganglion cells and nerve fibres. Ganglioneuromas may arise anywhere along the paravertebral sympathetic chain and occasionally from the adrenal medulla. However, the two most common are the retroperitoneum (32–52%) and the posterior mediastinum (39–43%). They may also occur in the cervical region. In the abdomen, ganglioneuromas occur either in the retroperitoneum (60%) or in the adrenal gland (40%). Although the tumours can occur at any age, the majority affect children and young adults (50%). Ganglioneuromas are often asymptomatic even when large. Otherwise, the clinical features include abdominal pain and a palpable abdominal mass. Hormonally active lesions secreting catecholamines, VIPs or androgens with associated hypertension, diarrhoea and virilisim are well documented. The CT or magnetic resonance scan shows a retroperitoneal or adrenal well-circumscribed tumour which may be oval or lobulated. These tumours tend to partially/completely surround major blood vessels but without narrowing the lumen. Additionally, ganglioneuromas exhibit heterogeneous contrast enhancement. One of the imaging characteristics of ganglioneuroma on MRI is curvilinear bands of low signal intensity on T_2 weighted images, giving the tumour a whorled appearance. The prognosis is excellent, and recurrence is rare after surgical resection.

Ganglioneuroblastomas and neuroblastomas

Ganglioneuroblastomas contain elements of both malignant neuroblastoma and benign ganglioneuroma. Histologically they behave as malignant lesions and contain a mixture of primitive neuroblasts and mature ganglion cells. The most common site of origin is the abdomen, followed by the mediastinum, neck and lower extremity. Macroscopically ganglioneuroblastomas are frequently encapsulated and contain granular calcification. Ganglioneuroblastomas usually occur in children aged 2–4 years with an equal sex incidence and are exceedingly rare after the age of 10. Their CT appearances vary, ranging from a predominantly solid mass to a predominantly cystic mass. Their prognosis and response to treatment are significantly better than neuroblastomas.

Neuroblastomas are malignant tumours consisting of primitive neuroblasts. They may arise anywhere within the sympathetic chain or adrenal medulla. On histology they are composed of small, dark neuroepithelial cells that may show glial or ganglionic differentiation and contain nests of primitive round cells with dark-staining nuclei and scanty cytoplasm. Clinically, they usually present during the first 10 years of life (80% occur in children <5 years) and are commoner in boys. Seventy per cent of neuroblastomas are intra-abdominal, with the majority originating from the adrenal gland. The remaining intra-abdominal abdominal/pelvic tumours originate from the paravertebral sympathetic chain or presacral area, and occasionally from the coeliac axis. Neuroblastomas tend to metastasize to bone, bone marrow, liver, lymph nodes and skin. Unfortunately, 70% of patients have metastatic disease at the time of diagnosis. Neuroblastomas may invade adjacent organs or encase blood vessels.

Both on plain radiography and CT, neuroblastomas usually contain areas of coarse calcification. Both CT and MRI scans are useful for preoperative assessment of extent of tumour and staging: involvement of retroperitoneal lymph nodes and hepatic deposits, tumour encasement of vessels and into the vertebral canal.

Tumours of the paraganglionic system

The paraganglionic system includes the adrenal medulla, the chemoreceptors (carotid and aortic bodies), vagal body, and the small groups of cells associated with the thoracic and intra-abdominal and retroperitoneal ganglia. Tumours that arise from the chromaffin cells of the adrenal medulla are called phaeochromocytomas or chromaffinomas, whereas those that occur in paraganglia at other sites are referred to as paragangliomas. On investigation, phaeochromocytomas are found to be the cause in 0.1–0.5% of patients with newly diagnosed hypertension. These tumours occur with equal frequency in men and women and usually present in the third and fourth decades. Phaeochromocytomas have been called '10% tumours' because 10% are bilateral, 10% are extra-adrenal (paragangliomas of the retroperitoneum, mediastinum or urinary bladder), 10% occur in children and 10% are malignant. When they occur in children, there is often a family history. Phaeochromocytomas may be associated with multiple endocrine neoplasia (MEN) 2a and 2b or 3 syndrome, von Hippel–Lindau syndrome and neurofibromatosis (NF-1, von Recklinghausen disease). MEN 2a syndrome consists of phaeochromocytoma, medullary carcinoma of the thyroid gland and parathyroid hyperplasia, whereas MEN 2b or 3 syndromes consist of phaeochromocytoma, medullary carcinoma of the thyroid gland, ganglioneuromatosis, and multiple mucosal neuromas. Phaeochromocytomas associated with MEN syndrome tend to be bilateral and are almost always intra-adrenal and usually benign.

The clinical manifestations of phaeochromocytoma result from excess and inappropriate secretion of catecholamine. The classic triad of headache, palpitation and excessive sweating is seen during the paroxysmal hypertensive crisis. Urinary metanephrine or vanillylmandelic acid levels are elevated in over 90% of patients. Confirmation by urinary 24 hour catecholamine excretion studies is followed by CT scanning of the adrenal gland. If no lesion is found in the adrenal, further CT studies are needed of the organ of Zuckerkandl (all chromaffin cell-bearing tissue along the lower abdominal aorta from the origin of the inferior mesenteric artery to the aortic bifurcation) and the urinary bladder (common site of paraganglioma). At CT, both phaeochromocytomas and paragangliomas are usually 3.0 cm or larger and demonstrate areas of necrosis or haemorrhage. Phaeochromocytomas are hypervascular and exhibit marked contrast enhancement after intravenous administration of contrast. If no lesion is found on CT and laboratory test results are positive on repeat testing, radionuclide imaging with meta-iodo-benzylguanidine (MIBG) may identify a small occult lesion. MRI can also be used to help locate a small paraganglioma.

Spontaneous rupture of a phaeochromocytoma or paraganglioma will produce a dramatic acute clinical presentation with severe hypovolaemic shock from massive retroperitoneal haemorrhage and severe back pain, not dissimilar to ruptured abdominal aortic aneurysm.

In general, paragangliomas exhibit a more aggressive behaviour than adrenal phaeochromocytomas. Thus 20–40% of paragangliomas metastasize, as opposed to 2–10% for adrenal phaeochromocytomas. Spread via the lymphatics and blood is to regional lymph nodes, bone, liver and lung.

Polyposis syndromes involving the small bowel

These syndromes are characterized by a common pathological feature – the presence of multiple polyps which may affect all regions of the gastrointestinal tract. Individually, each syndrome is distinguished by its aetiology (genetic or otherwise), the microscopic appearance of the polyps and the risk of invasive malignancy. The four polyposis syndromes which involve the small intestine are:

- familial adenomatous polyposis (FAP)
- Peutz–Jeghers syndrome (PJS)
- generalized juvenile polyposis
- Cronkhite–Canada syndrome (CCS).

Familial adenomatous polyposis

This hereditary autosomal dominant condition is characterized by the development of multiple adenomatous polyps in the colon and which inevitably progress to invasive colonic cancer (see Chapter 30). Additionally, however, some 80% of patients with FAP also have adenomas involving the small bowel and, like the colonic polyps, these are also premalignant. They tend to occur mainly in the duodenum in the region of the ampulla of Vater (second part). In patients with FAP who have had their colon removed, the most common cause of cancer death is from periampullary cancer. When found, large duodenal adenomas in FAP patients should be removed surgically or endoscopically. There is some evidence that certain anti-inflammatory agents may reduce the growth of some of these small bowel adenomas in FAP colectomized patients. However, surveillance by flexible endoscopy or CE is advisable in all these patients every 2 years.

Peutz–Jeghers syndrome

This is an autosomal dominant disorder consisting of mucocutaneous pigmentation (on the face, lips and buccal mucosa) and benign gastrointestinal hamartomas. The polyps are found in the small bowel in up to 90% of affected individuals but the polyps can also be in the stomach and colon. The hereditary defect is associated with mutation of the *LKB1* gene (19p2, 3). Complications are common and include colicky abdominal pain, gastrointestinal bleeding, intestinal obstruction and the development of gastrointestinal and various extraintestinal cancers.

Aetiology and pathology

PJS is characterized by intestinal hamartomatous polyps in association with mucocutaneous melanocytic macules. The cause of PJS appears to be a germline mutation of the serine threonine kinase (*STK11)* gene, which is present in the majority of patients. This protein is regulated by phosphorylation by cAMP-dependent protein kinase A. *STK11/LKB1* (serine/threonine kinase 11) is a tumour suppressor gene (overexpression induces growth arrest of cells at the G1 phase of the cell cycle) located on band 19p13.3. Somatic inactivation of the unaffected allele of *STK11* is frequently present in polyps and cancers from patients with PJS. The variable penetrance of the gene accounts for the varied phenotypic manifestations encountered clinically in patients with PJS: different number and location of polyps and different pigmentation patterns. Although other genes may be involved in the aetiology of PJS, the current evidence is that *LKB1* defects are the major cause of PJS, although other mechanisms may be involved. Adnab-9, which is a premalignant marker found in Paneth cells, is more common in PJS than in familial juvenile polyposis and hyperplastic polyps. Thus Adnab-9 labelling may identify polyps at higher risk of malignant degeneration.

Histologically, the polyps of PJS exhibit a distinctive frond-like appearance with a stromal/smooth muscle core covered by acinar glands and normal mucosa without any nuclear atypia. The number, as well as the size and the location, of the polyps vary from patient to patient. Isolated melanotic mucocutaneous pigmentation without gastrointestinal polyps has also been described and is explained by the genetic variability of the syndrome.

The principal cause of morbidity in PJS results from the intestinal location of the polyps (small intestine, colon, stomach). Complications, which usually occur in the second and third decades, include small bowel obstruction and intussusception (43%), gastrointestinal bleeding (14%) and rectal prolapse of a colonic polyp. Almost 50% of patients with PJS develop and die from cancer by the age of 60, with the mean age at first diagnosis of cancer being 43 years. The cumulative risk for the development of specific cancers is as follows: breast 54%; colon 39%; pancreas 36%; stomach 29%; ovary 21%; lung 15%; small intestine 13%; cervix 10%; testes 9%; and uterus 9%. The statistically significant elevated relative risk (RR) and confidence intervals (CI) for cancer in the various organs relative to the general population is as follows:

- all cancers (RR 15.2; CI 12–19)
- oesophagus (RR 57; CI 2.5–557)
- stomach (RR 96; CI 96–368)
- small intestine (RR 520; CI 220–1306)
- colon (RR 84; CI 47–137)
- pancreas (RR 132; CI 44–261)
- lung (RR 17; CI 5.4–39)
- breast (RR 15.2; CI 7.6–27)
- uterus (RR 16.0; CI: 1.9–56)
- ovary (RR 27; CI 7.3–68).

Clinical features

PJS is rare, with a reported frequency of three per 300 000 population, i.e. approximately one-tenth that of FAP. The sex incidence is equal. The average age at diagnosis is 23–26 years. The clinical features of PJS include a family history, repeated bouts of abdominal pain and obscure intestinal bleeding in patients younger than 25 years, rectal prolapse, menstrual irregularities in females (due to hyperostrogenism), gynaecomastia in males (oestrogens production from Sertoli cell testicular tumours) and precocious puberty. Small bowel obstruction develops in about 50% of patients. The obstruction may be caused by the polyp itself or result from intussusception initiated by a polyp.

Physical examination shows mucocutaneous pigmentation and melanin spots are present in more than 95% of patients. The pigmented lesions are present in the first years of life and may fade at puberty with the exception of lesions on the buccal mucosa which persist in all cases. The pigmented lesions consist of small, flat, brown or dark blue spots/macules (1–5 mm) resembling freckles, most commonly in the circumoral and perinasal area. The pigmentation may also be present in the fingers and the toes, on the dorsal and volar aspects of the hands and the feet and around the anus and genitalia. A rectal mass (rectal polyp) or a large polyp can prolapse outside the anus.

Treatment

There is no specific medical therapy. PJS should be diagnosed as early as possible and genetic counselling provided. Screening for both intestinal cancers and extraintestinal cancers should be implemented once the diagnosis is established. This includes an annual physical examination of the breasts, abdomen, pelvis and testes. Surveillance for cancer has been suggested with small bowel enema every 2 years (and, more recently, CE), upper gastrointestinal endoscopy, colonoscopy every 2 years, annual ultrasound scanning of the pancreas, pelvis (in women) together with cervical smears (Papanicolaou) and examination of the testes in males. Mammography is advised at 25, 30, 35 and 38 years and thereafter every 2 years until age 50. In addition, all haemorrhagic or polyps >5 mm should be removed by endoscopic polypectomy

Surgical treatment is reserved for complications: intestinal obstruction (including intussusceptions), resection of large polyps, persistent gastrointestinal bleeding and the development of invasive cancers.

Generalized juvenile polyposis

Generalized juvenile polyposis can be inherited or sporadic. The polyps can arise anywhere in the intestinal tract. They have a distinctive microscopic appearance and can bleed or cause intestinal obstruction when large. The risk of invasive gastrointestinal cancer is however small. Juvenile gastrointestinal polyposis is a syndrome described in infants who present with hamartomatous polyposis, macrocephaly, alopecia, nail dystrophy, clubbing of fingers and toes, hypotonia, hepatosplenomegaly, anaemia and hypoproteinaemia due to protein-losing enteropathy. Although some authors consider this syndrome a form of CCS, others think it is a different entity. The prognosis is usually very poor.

Cronkhite–Canada syndrome

CCS, first described by Cronkite and Canada in 1955, is characterized by intestinal polyps, pigmentation, diarrhoea and protein-losing enteropathy, which may lead to severe weight loss from malnutrition, dystrophic changes in the fingernail, alopecia and abdominal pain. The polyps are most commonly found in the small intestine and usually do not become malignant. CCS is not familial. The aetiology is uncertain, although an autoimmune process is thought to be responsible. Thus many patients have elevated antinuclear antibody (ANA) and immunoglobulin (Ig)G4 levels. Other suggested aetiological factors include mental stress and physical fatigue. The disease is very common in the Japanese, who form 75% of the reported cases. The estimated incidence of CCS is one per million. The mean age of onset is in the fifth to sixth decade, with a slight male predominance. Endoscopically, the gastric mucosa may be thickened with hypertrophic gastric folds mimicking Ménétrier disease or be atrophic with polypoid lesions. The colonic polyps in CCS are usually sessile with a ‘strawberry-like’ appearance.

Pathophysiology

The gastrointestinal changes consist of a generalized hamartomatous polyposis with abnormal intervening mucosa. The prevalence of gastrointestinal malignancy in patients with CCS is about 10%. The polyps are part of a generalized gastrointestinal mucosal disturbance that results in malabsorption and protein-losing enteropathy leading to severe diarrhoea, abdominal pain and profound malnutrition. Dermatological manifestations accompany the gastrointestinal symptoms. The vast majority of affected individuals have alopecia, alterations in the nail beds and skin pigmentation. Although the published data support the view that ectodermal signs are consequent on the severe malnutrition, this is by no means certain, especially as the ectodermal changes may precede the gastrointestinal symptoms and signs. The aetiology is unknown. Japanese reports have suggested mental stress, family problems and physical fatigue as the most frequent precipitating factors. Patients with CCS often have coexisting autoimmune disorders, e.g. type 1 diabetes, hypothyroidism, membranous glomerulopathy, and high titres of ANA.

A recent report described the infiltration by IgG4 producing plasma cells in 50% of patients with CCS, suggesting it belongs to the group of IgG4-related autoimmune diseases which include autoimmune pancreatitis, sclerosing cholangitis and peritoneal fibrosis.

Clinical features

The disease is commoner in males and in Japan; the male-to-female ratio is 2:1. The age distribution of published cases ranges from 31 to 85 years, with onset of disease occurring most commonly in patients aged 50–60 years. The initial symptom is either diarrhoea or hypogeusia (taste disturbances). Clinically, the disease usually exhibits a rapid progression over several months in the severity of the gastrointestinal symptoms leading to substantial weight loss, malnutrition and peripheral oedema. Ectodermal changes are usually observed several weeks or months after gastrointestinal symptoms begin. Aside from the consequences of malnutrition, which can become life threatening, most of the complications encountered are manifestations of the polyposis. Some patients with CCS present initially with other gastrointestinal disorders including *Helicobacter pylori* gastritis, eosinophilic gastroenteritis and intestinal candidiasis.

The diarrhoea, which is present in 90% of patients, is multifactorial. The intestinal mucosal glands secrete a copious amount of protein-rich fluid. Additionally, diseased small bowel mucosa is unable to digest and absorb disaccharides and triglycerides. Additionally, the normal gut microflora is replaced by bacterial overgrowth. Patients typically have five to seven loose watery bowel movements per day with volumes averaging 4–6 L. Abdominal pain, anorexia, vomiting, weakness and weight loss accompany the diarrhoea and steatorrhoea. Weakness is related to protein calorie malnutrition with muscle wasting, dehydration and faecal electrolyte losses (calcium, magnesium, potassium and zinc).

The morbidity of CCS stems from malabsorption (vitamins, carbohydrates, lipids, vitamins, fluid and electrolyte, and enteric losses of serum proteins), resulting in oedema, congestive heart failure and immunodeficiency. Iron-deficiency anaemia results from chronic blood loss with iron and folate deficiency. Rarely, massive bleeding may occur and prove fatal. The early estimates of a 50% 5 year death rate in patients with CCS have been significantly reduced by improved supportive therapy.

Physical examination reveals characteristic ectodermal changes starting several weeks or months after the initial gastrointestinal symptoms in almost all patients. Most patients have two or more of the cutaneous triad that consists of alopecia (initially patchy but progresses to complete hair loss), nail changes in fingers and toes (onychodystrophy), consisting of thinning, splitting and partial separation of nail from nail bed, and hyperpigmentation, consisting of brownish macules or plaques most commonly on the hands and arms but may occur elsewhere and be widespread.

Treatment

This is essentially supportive with fluid and electrolyte replacement and restoration/maintenance of macronutrient and micronutrient requirements. Mild malnutrition is treated with replacement of fluids and electrolytes using elemental diets, but severe cases of malnutrition require complete bowel rest and total parenteral hyperalimentation.

Antibiotics are administered orally for bacterial overgrowth. Systemic infections are treated with broad-spectrum antibiotics administered parenterally. Hydrocortisone sodium succinate is the most commonly used corticosteroid to decrease inflammation by reversing capillary permeability and suppression of polymorphonuclear cell activity. Other drugs used include proton pump inhibitors for gastric acid suppression, alone or in combination with appropriate antibiotics to eradicate *H. pylori* infection, cromolyn and zinc/vitamin supplements. Sulindac (a non-steroidal anti-inflammatory drug) does not induce regression of CCS polyps. Surgical treatment is reserved for the treatment of complications, i.e. bleeding, obstruction (usually from intussusception) and the development of invasive malignancy.

Malignant small bowel tumours

Malignant small bowel tumours are extremely rare, accounting for less than 5% of all malignant gastrointestinal neoplasms. Several factors have been suggested to explain the rarity of both benign and malignant small bowel tumours. The first explanation relates to the rapid small bowel transit time, which limits contact time of any carcinogen with the small bowel mucosa. The second concerns the large volume and fluid nature of the small bowel content, which dilutes any intraluminal irritant/carcinogen. The third protective factor is attributed to the high pH of the intestinal content associated with the low bacterial colony count of the small bowel. Additionally, the small bowel contains high levels of the enzyme benzpyrene hydroxylase, which destroys potential carcinogens.

The downside to the rare incidence of malignant small bowel tumours is their asymptomatic early phase, such that their clinical presentation is usually with advanced disease with tumour spread beyond the confines of the bowel wall by the time of diagnosis. There are four important malignant small bowel tumours:

- adenocarcinoma (40%)
- carcinoid tumours (30%)
- lymphomas (25%)
- GISTs (5%), discussed previously.

In many cases, diagnosis is usually made at laparotomy for small bowel obstruction. However, increasingly early-stage tumours are being diagnosed by CE. The other diagnostic modalities which may be useful include small bowel follow-through, small bowel enema and small bowel enteroscopy, including single-balloon endoscopy and DBe, mesenteric angiography and CT. In some cases laparoscopic examination of small intestinal loops may be useful and enables biopsy of serosal and mesenteric masses.

Adenocarcinoma

Adenocarcinomas of the small intestine are rare, being 50–60 times less common than colonic adenocarcinoma. They are mucus-secreting tumours and are commonest in the proximal part of the small intestine (duodenum 40%, jejunum 40%, ileum 20%). The majority of duodenal carcinomas are found in the periampullary region and in the third part of the duodenum. Intestinal adenocarcinomas spread primarily to the regional lymph nodes, liver and peritoneal cavity. The prognosis is poor largely because of late presentation, resulting in an overall 5 year survival of 15%. Surgery is the mainstay of treatment as these tumours are largely unresponsive to chemotherapy and radiotherapy.

Aetiology and pathology

The aetiological factors include genetic and environmental risk factors in addition to specific small bowel disorders which predispose to the disease. The genetic conditions include FAP as these patients in addition to colon cancer are at risk of duodenal adenocarcinoma (RR >300). Aside from colorectal carcinoma, patients with hereditary non-polyposis colorectal cancer (HN-PCRC) can also develop other cancers, including small bowel adenocarcinoma (>100 times that of the general population). In these patients, the cancers are distributed evenly throughout the small bowel, occur at younger age and have a better prognosis than sporadic small bowel cancers. The most commonly mutated genes in patients with HN-PCRC are *HMLH1* and *HMSH2*, which are involved in DNA mismatch repair.

Environmental risk factors include high consumption of animal fat, red meat and salt-cured or smoked foods and smoking. Other than FAP and HN-PCRC, the predisposing disorders/conditions for small bowel cancer include the following.

- Crohn's disease: RR variously estimated between 15 and >100. Adenocarcinoma occurs late, usually 20 or more years after the onset of Crohn's disease, and the majority develop in the ileum, reflecting the distribution of the disease.
- Coeliac disease: carries a RR of 300 for the development of lymphoma and 65 for the development of small bowel adenocarcinoma.

- PJS: associated with an 18-fold increase in the incidence compared to the general population.
- Ileostomy sites.
- Surgically bypassed duodenum.
- Jejunal limb of a Roux-en-Y oesophagojejunostomy.
- Ileal segment of a defunctionalized ileocystoplasty.

Pathologically, the tumours are graded as grade I (well differentiated), grade II (moderately differentiated) and grade III (poorly differentiated). Small bowel adenocarcinomas are more common in the upper reaches of the small bowel with 40–50% affecting the duodenum and arising from the mucosa in the region of the ampulla of Vater or from the third part. Periampullary tumours usually form papillary growths. More distal lesions may be polypoid, annular or fungating and often present with acute intestinal obstruction. Small bowel adenocarcinomas produce mucin and, like large bowel cancers, exhibit carcinoembryonic antigen (CEA) reactivity and express COX-2, sPLA2¢ and cPLA2. Endocrine cells (chromogranin and 5-HT) are also present scattered among the adenocarcinoma cells, particularly in ileal tumours. Rarely, these endocrine cells exhibit immunoreactivity for various peptide hormones: somatostatin, YY peptide, neurotensin, glucagon, etc.

Clinical features

In general, men have higher rates of all types of small bowel cancer than females (M/F ratio of 1.4:1.09). The prevalence increases with advancing years, with a mean age at diagnosis averaging 60 years. Small bowel cancer is asymptomatic in its early stages, and hence the majority present with advanced disease. The common presenting symptoms include abdominal discomfort, which is usually postprandial and colicky in nature, nausea and vomiting, particularly in patients with duodenal carcinomas and weight loss. Intestinal obstruction occurs in 50% of patients with the diagnosis being made during emergency surgery. Adenocarcinomas are more frequently associated with pain and obstruction than small bowel sarcomas (including GISTs), which tend to present more commonly with gastrointestinal bleeding. When it occurs, gastrointestinal bleeding may be occult leading to iron-deficiency anaemia but may also be overt leading to melaena or frank rectal bleeding. Intestinal obstruction usually indicates advanced disease, although a relatively small polypoid tumour may lead to intussusception. Patients with periampullary duodenal carcinomas usually present with obstructive jaundice.

The investigations used to establish the diagnosis depend on the clinical presentation. In patients presenting with bleeding, upper gastrointestinal endoscopy and colonoscopy are usually performed in the first instance. Colonoscopy with retrograde ileoscopy may be useful in identifying distal ileal tumours. In the absence of bleeding or acute presentation with small bowel obstruction, small bowel follow-through or small bowel enema (enteroclysis) have been traditionally used for establishing the diagnosis, but are being replaced in many centres with small bowel endoscopy and balloon endoscopy. Although small bowel enteroclysis with the double contrast technique has a sensitivity of 95%, it is difficult to perform and unpleasant for the patient. Plain abdominal X-ray films are indicated in patients presenting with partial or complete small bowel obstruction.

The American Joint Committee on Cancer staging system is generally used for the staging of small bowel malignant tumours.

Primary tumour (T)

- TX: Primary cannot be assessed
- T0: No evidence of primary tumour
- Tis: Carcinoma *in situ*
- T1: Tumour invades the lamina propria or submucosa
- T2: Tumour invades the muscularis propria
- T3: Tumour invades through the muscularis propria into subserosa or into non-peritonealized perimuscular tissue (mesentery or retroperitoneum), with extension of less than 2.0 cm
- T4: Tumour penetrates the visceral peritoneum or directly invades other organs or structures

Regional lymph nodes (N)

- NX: Regional lymph nodes cannot be assessed
- N0: No regional lymph node metastasis
- N1: Regional lymph node metastasis

Distant metastases (M)

- MX: Presence of distant metastasis cannot be assessed
- M0: No distant metastasis
- M1: Distant metastasis

Stage grouping

- Stage 0: is, N0, M0
- Stage I: T1–2, N0, M0
- Stage II: T3–4, N0, M0
- Stage III: Any T, N1, M0
- Stage IV: Any T, any N, M1

Treatment

Surgical resection provides the only hope of cure for patients with small bowel adenocarcinomas. Resection with curative intent or palliative is possible in 70% of cases with the remainder having unresectable disease at the time of presentation because of extensive local disease or metastases to regional lymph nodes, liver and peritoneal serosal surfaces. In patients deemed potentially curable, a wide radical (5.0 cm margins) segmental excision with regional lymphadenectomy is performed. In patients with duodenal adenocarcinoma, this necessitates radical pancreaticoduodenectomy. Tumours close to the ileocaecal valve may require a right hemicolectomy. In published series, 5 year survival rates of 40–60% have been reported with radical surgery. Surgical excision may also be indicated for palliation in patients with symptomatic advanced disease causing intestinal obstruction.

Although various chemotherapy regimens have been used in patients with incurable disease most commonly with 5-fluorouracil (5FU) based regimens, FOLFOX 4 (infusional 5FU, oxaliplatin and leucovorin) and irinotecan, the benefits in terms of response rates and survival benefit have been poor. Similarly the addition of adjuvant chemotherapy after surgery does not improve outcome of patients after surgical resection.

Carcinoid tumours

These are neuroendocrine tumours (NETs) which include *carcinoid tumours* and *pancreatic endocrine tumours*. Carcinoid tumours account for the largest number of NETs and can produce the carcinoid syndrome. NETs originate from enterchromaffin-like cells, which normally control acid secretion by releasing histamine and which also regulate the *motility of the gastrointestinal tract.* The commonest sites of carcinoids of the gastrointestinal tract are the appendix, small intestine and the rectum in that order (see The vermiform appendix section for further discussion of carcinoid tumour of the appendix).

Pathology

Carcinoid tumours arise from the Kulchitsky cells of the crypts of Lieberkuhn. These neuroendocrine cells are also referred to as 'enterochromaffin cells' as they stain with potassium chromate. The aetiology of carcinoid tumours is not known, but genetic abnormalities are suspected. Multiple carcinoid tumours occurring in association with MEN 1 are well documented. Carcinoid tumours which occur predominantly in the gastrointestinal tract are classified into three groups.

- *Foregut tumours* may be sporadic primary tumours or secondary to achlorhydria. Sporadic foregut tumours occur in the lung, bronchus, stomach, proximal duodenum and pancreas. These tumours consist of regularly shaped cells which assume a trabecular arrangement and contain round granules. They exhibit an argyrophilic reaction, which indicates that the cells can only be stained with metallic silver in the presence of a reducing agent.
- *Midgut tumours* arise in the second portion of the duodenum, jejunum, ileum and the right colon. They account for 60–80% of all carcinoid tumours (especially in the appendix and distal ileum) in adults but may also be encountered in children commonly as apical lesions of the appendix, often discovered as an incidental finding in appendicectomy specimens. The tumour cells are pleomorphic and are arranged in nests separated by connective tissue. These cells have both argentaffin (can be stained directly with metallic silver without reducing agent) and argyrophilic staining properties.
- *Hindgut tumours* occur in the left colon and rectum. The cells here are arranged in a trabecular pattern containing round granules but do not stain with silver.

Foregut tumours produce 5-HTP, 5-HT (serotonin), histamine and substance P. Midgut tumours produce 5-HT, kallikrein and prostaglandins. Tumours of the hindgut do not usually secrete active peptides. In addition, midgut and foregut tumours may contain and secrete a variety of hormones including insulin, somatastatin, adrenocorticotrophic hormone (ACTH), gastrin, antidiuretic hormone, parathormone, glucagon, VIP, calcitonin, β-melanocyte-stimulating hormone, cholecystokinin and growth hormone. The cytology and histology cannot differentiate between benign and malignant carcinoid tumours. In general, tumours smaller than 1.0 cm are rarely malignant. Tumours between 1.0 and 1.9 cm may be malignant, and tumours larger than 2.0 cm are usually invasive and may exhibit metastatic spread. These more aggressive lesions invade the bowel wall mesentery, parietal peritoneum and adjacent organs. Metastatic spread involves the regional lymph nodes, peritoneum and liver in particular, but other sites such as the lung, spleen, ovaries and the bones may be affected.

Carcinoid tumours of the duodenum are rare. They tend to be similar in nature to other gastrointestinal carcinoids but there is an unusual type known as the carcinoid islet cell tumour which has the functional, morphological and histochemical features of both foregut carcinoids and pancreatic islet cell tumours (APUdomas). In addition to serotonin, the tumour may also produce and secrete a variety of peptide hormones such as gastrin, insulin, parathyroid hormone and catecholamines causing bizarre clinical manifestations.

Carcinoid tumours of the small intestine are most commonly encountered in the ileum. The majority are malignant, 40% are multiple and in 30% there is an associated malignant neoplasm usually an adenocarcinoma. In the gastrointestinal tract, carcinoids develop deep in the mucosa and with growth they extend into the underlying submucosa, thus forming small firm yellowish grey nodules, which bulge into the intestinal lumen. The yellow colour results from the accumulation of cholesterol and lipids. Carcinoid tumours can have a polypoid appearance and may become ulcerated. With increased invasive growth, they eventually infiltrate the muscularis propria to reach the serosa and the mesentery when they are invariably accompanied by metastatic spread.

Three types of carcinoid tumours are differentiated on histology: (1) classical, (2) tubular adenocarcinoid and (3) goblet cell carcinoids (GCCs). The classical type forms solid nests of small uniform cells with occasional acinar or rosette formation, infrequent mitosis and retraction of tumour periphery from stroma, with some cells being located within intra-appendiceal nerves (may be related to histogenesis). They usually invade the muscularis and may spread to the serosal surface. Ultrastructurally the component cells contain pleomorphic dense-core secretory granules. The tubular (adenocarcinoid) type of carcinoid is characterized by glandular architecture without solid nests and lack mitoses and atypia. The cytoplasmic granules are large and acidophilic. The GCC exhibits a predominantly submucosal growth with frequent extension into muscle and serosa, the mucosa being spared. Some cells may be lysozyme-positive cells with features of Paneth cells. Ultrastructurally the cells contain mucin droplets and neurosecretory type granules at times in the same cell (amphicrine cells). GCCs are discussed further under Carcinoid tumours of the appendix.

In the midgut carcinoid lesions may be associated with diffuse thickening of the bowel wall. Carcinoid tumours commonly affect the tip of the appendix but they often invade the wall of the appendix and spread to the regional lymph nodes, and about 75% of patients have evidence of peritoneal involvement. The high potential for metastasis by carcinoid tumours is thought to be caused by the production and release of β-catenin. The size of the tumour correlates with the stage of the disease; tumours smaller than 1.5 cm are rarely accompanied by distant metastases or recurrence after surgical excision.

Carcinoid tumours can be associated with concentric and elastic vascular sclerosis, which may result in obliteration of the vascular lumen and ischaemia. Fibroblastic proliferation is caused by stimulation of fibroblast cells by growth factor produced by the tumour. Carcinoids often express somatostatin receptors which are used for imaging and treatment of the disease.

Clinical features

Carcinoid tumours are rare. In the USA, the estimated incidence is 1.5 cases per 100 000 individuals. The overall age-standardized incidence rate for male and female populations in England is 0.71 and 0.87, respectively. In Scotland, the incidence rate is 1.17 and 1.36 per 100 000 for males and females, respectively. The average age of reported cases of small intestinal carcinoid is between 45 and 55 years. Duodenal carcinoids may present with vomiting due to obstruction or as an endocrine syndrome. Carcinoids of the jejunum or ileum present with diarrhoea, intestinal obstruction, palpable abdominal mass and, much less commonly, massive gastrointestinal haemorrhage or intestinal infarction.

An early symptom of carcinoid tumours, especially in patients with metastases from midgut tumours, is cutaneous flushing. This typically affects the head and neck. Striking colour changes range from pallor or erythema to cyanosis. Episodes are often associated with an unpleasant warm feeling, itching, palpitations, salivation, lacrimation and diarrhoea. Exercise, stress or certain foods (e.g. cheese) may trigger an attack, although the flushing episodes may also be spontaneous. With time, the duration of the flushing episodes increases and may last hours, and, indeed, some patients develop a constant red/cyanotic discoloration.

Diarrhoea and malabsorption occur in the majority of patients. The stools are frothy and bulky (steatorrhoea). The diarrhoea may or may not be associated with abdominal pain, flushing, and cramps. Wheezing caused by bronchospasm may occur and be accompanied by dyspnoea. Relatively uncommon complaints include joint pain, arthritis, lacrimation, confusion and changes in mental status, ophthalmological findings associated with flushing or secondary to vascular occlusion, retroperitoneal fibrosis, obstruction of ureter, intra-abdominal fibrosis, and male sexual dysfunction. Carcinoid tumours can occur in association with familial or genetic disorders, such as MEN 1 and PJS.

Carcinoid syndrome

The carcinoid syndrome is very rare, produced by less than 10% of all carcinoid tumours and in the majority of cases the primary tumour originates in the small intestine. The syndrome is invariably associated with the presence of extensive hepatic involvement by tumour and the clinical manifestations are the result of inappropriate secretion of 5-HT, 5-HTP, kallikrein, histamine, prostaglandin and ACTH, etc. Production of VIP may produce symptoms similar to those of neuroblastoma and ectopic secretion may produce Cushing syndrome induced by foregut carcinoid tumours. Carcinoid crisis can occur spontaneously or as a response to stress, e.g. anaesthesia or chemotherapy. Symptoms may include intense flushing, diarrhoea, abdominal pain, tachycardia, hypertension or hypotension, altered mental status and coma. This condition can be life threatening, but treatment with somatostatin analogue SMS-201-995 has improved the outcome of patients with carcinoid crisis.

In the absence of carcinoid crises, the characteristic symptoms of the carcinoid syndrome include various types of flushing episodes, intestinal colic and diarrhoea, bronchospasm and hypoproteinaemia. The cutaneous flushing episodes affect the upper part of the body and are accompanied by sweating, itching, oedema, palpitations and hypotension. In time, patients may develop other manifestations including cardiac lesions, pellagra-like skin lesions (photosensitive dermatitis), neurological signs, peptic ulceration and neuralgia. Cardiac lesions are present in up to 60% of patients and are the result of endocardial fibrosis, which tends to involve the right heart, especially the ventricular aspect of the tricuspid valve and associated chordate. Stenosis from fibrosis of the pulmonic valve is less common but involvement of the mitral valve is rare. These cardiac lesions may lead to heart failure.

Biochemical confirmation of the diagnosis of the carcinoid syndrome is achieved by the determination of the urinary metabolite of 5-HT and 5-HTP, 5-hydroxyindole acetic acid. The intravenous administration of 2 μg of epinephrine (adrenaline) results in a typical attack of flushing within 2 minutes of the injection, but this test is no longer used.

Treatment of carcinoid tumours and carcinoid syndrome

Whenever possible, complete surgical excision is the treatment of choice. In small bowel carcinoids, block resection of the tumour with regional lymph nodes is needed. In rectal tumours, endoscopic resection is sufficient for tumours measuring less than 1.0 cm limited to the mucosa, otherwise more extensive sphincter-saving resections are needed. The treatment of appendiceal carcinoids is discussed in the section Carcinoid tumours of the appendix.

For secondary hepatic tumours, surgical ligation of the hepatic artery can potentially deprive the blood supply to the tumour cells and cause necrosis while preserving most of the normal liver cells. However, with time, new blood vessels develop and restore circulation such that the clinical benefit is short lived, and, for this reason, this therapeutic approach is seldom used nowadays. If hepatic metastases are resectable, hepatic surgical resection is the preferred treatment. Resection usually involves hepatic plurisegmentectomies. Intra-arterial infusion of chemotherapeutic agents with chemoembolization of the hepatic artery may also provide effective, albeit short term, relief of symptoms due to hepatic metastasis. Both resection and embolization should be covered with antiserotonin therapy. Arterial embolization is performed percutaneously using a selective angiographic technique whereby the arteries feeding the metastases are blocked by gelatin sponge or human dura mater delivered in an antibiotic-containing solution and followed by steel coils. In addition to antiserotonin therapy these patients must be covered by systemic antibiotics and steroids.

Lesser degrees of palliation can be obtained by chemotherapy. The drugs which have been shown to be useful are doxorubicin, 5FU, dacarbazine, actinomycin D, cisplatin, etoposide, streptozotocin, interferon-α and somatostatin analogues with a radioactive load used singly or in combination. The best results have been reported with prolonged infusion chemotherapy, either through the hepatic artery or a tributary of the portal vein. Prolonged access to the hepatic arterial tree is possible with the use of implantable subcutaneous chambers which have a silicon diaphragm allowing intermittent and prolonged infusions (Figure 29.17).

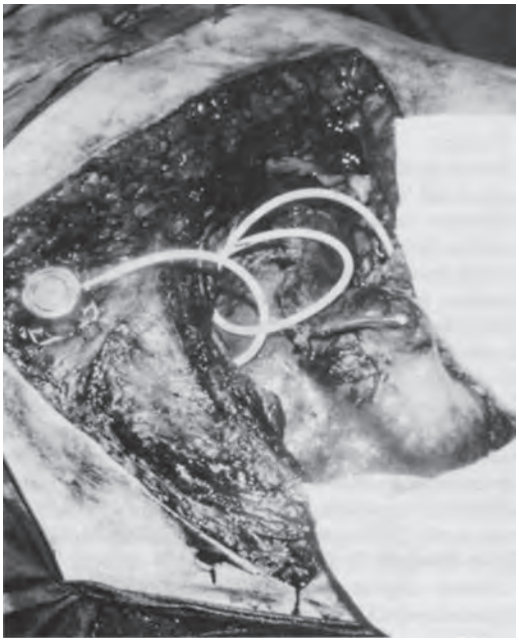


Figure 29.17 Port-A-Cath implantable venous/arterial access system. In this particular patient, the technique has been used for prolonged hepatic arterial infusion chemotherapy for hepatic secondary deposits. In between treatments, the system is left primed with heparinized saline.

Totally implantable infusion pumps are available, but they are expensive and do not have any special advantage over implantable access systems connected to small portable battery power infusion pumps. In the case of metastatic carcinoids, pretreatment with antiserotonin therapy is necessary to prevent a carcinoid crisis during chemotherapy. Antiserotonin therapy can be undertaken with agents that either reduce the production of 5-HT or antagonize its effects (Table 29.3)

Octreotide (synthetic somatostatin analogue) is highly effective in controlling symptoms but without imparting a survival advantage. The best results have been reported with continuous infusions which substantially reduce the plasma insulin-like growth factor I and growth hormone-releasing hormone. The long-acting somatostatin analogue which is administered monthly has removed the need for daily injections/infusions. The adverse effects of prolonged somatostatin therapy include gall stones and steatorrhoea, which may sometimes require pancreatic enzyme replacement. *In situ* targeted therapy with somatostatin analogues (octreotide attached to radioactive yttrium-90 or indium-111) provides promise and is currently undergoing clinical trials in patients with unresectable tumours.

Table 29.3 Antiserotonin drugs and their effects

Parachlorophenylalanin	Relieves diarrhoea
Phenoxybenzamine	Administered to achieve alpha blockade. It may relieve attacks precipitated by emotion, diet, exercise and alcohol
Chlorpromazine	Has antikinin effects and may control flushing
Methysergide maleate	Most potent antagonist of 5-HT. May relieve flushing, diarrhoea and bronchospasm
Cyproheptadine	Less potent antiserotonin agent. May relieve diarrhoea and less frequently reduce intensity of flushing
Prednisolone	May relieve facial oedema, diarrhoea and the flushing symptoms of bronchial carcinoids but not when the symptoms are caused by gastrointestinal tumours
Long-acting somatostatin analogue	Abolishes flushes due to gastrointestinal carcinoids and diarrhoea
Ketanserin	May reduce diarrhoea
Calcitonin	Same effects as somatostatin

5-HT, 5-hydroxytryptamine.

Small bowel lymphoma

Aetiology and pathology

An important distinction must be made between primary intestinal non-Hodgkin's lymphomas (NHLs) and the much more common secondary involvement of the subdiaphragmatic region (abdomen and its viscera) by systemic Hodgkin's nodal or extranodal disease. The agreed criteria for the diagnosis of primary intestinal NHLs are:

- no palpable superficial lymphadenopathy
- no mediastinal lymphadenopathy on chest radiograph/CT
- normal white cell count, normal bone marrow examination
- patients present with gastrointestinal symptoms
- bowel lesion predominates at laparotomy and only regional lymph nodes are involved
- liver and spleen are not involved.

On the basis of these diagnostic criteria, it is likely that in Western countries primary lymphoma accounts for less than 30% of all patients with intestinal involvement by lymphomatous disease. As primary Hodgkin's disease is extremely rare in the gut (1%), for practical purposes all primary intestinal lymphomas are NHLs.

The sites of origin of primary gastrointestinal NHLs are:

- stomach 50–75%
- small bowel 9.0%
- ileocaecal region 7.0%
- involvement of more than one gastrointestinal site 6.0%
- other sites (duodenum, colon, rectum) 3.0%.

Although the exact aetiology of the majority of cases of NHLs of the bowel is unknown, certain predisposing disorders are well documented. These include coeliac disease, ulcerative colitis and Crohn's disease, AIDS and other immunodeficiency states, e.g. transplant patients, chronic lymphatic leukaemia and long-term cyclophosphamide treatment for other forms of malignancy and various infections.

All types of small bowel NHLs are commonest in young adult males (aged 17–53 years) of low socioeconomic class mainly in developing Middle Eastern countries, north and south Africa and the Far East. Immunoproliferative small intestinal disease (IPSID) or Mediterranean lymphoma affects older children and young adults (aged 10–35 years), again in Middle Eastern/Mediterranean countries, although population studies from the Middle East indicate a decline in the incidence of IPSID over the last three decades.

Small bowel NHLs are classified as -cell or T-cell lymphomas, each of which can be further subdivided into low, intermediate and high grade. The majority are -cell lymphomas arising from the centrocyte-like cells of the mucosa-associated lymphoid tissue and referred to as MALT lymphomas. The other common types of intestinal NHLs include the centrocytic lymphomas, Mediterranean lymphoma, Burkitt-type lymphoma and the polymorphic T-cell lymphoma.

The low tendency of small bowel NHLs to disseminate outside the gastrointestinal tract may be due to the expression of specific adhesion molecules and/or dependence on local stimuli such as antigens or chemokines. Mucosal lymphocytes strongly express the $\alpha4\beta7$ integrin, whereas its ligand, MAdCAM-1, is selectively expressed on mucosal endothelium. This explains the established expression of the mucosal homing receptor $\alpha4\beta7$ by mucosa-associated B-NHLs (e.g. low-grade MALT lymphomas and mantle zone lymphomas). Additionally, intestinal epithelial cells produce several chemokines, i.e. CLL25 (TECK), CCL5 (RANTES) etc. The receptors for these chemokines are expressed by $\alpha4\beta7+$ T-lymphocytes present in the lamina propria and in the epithelium. The chemokine CLL25 (TECK) has been shown to attract IgA-secreting cells to the intestine. The chemokine receptor CXCR3 is expressed by a small subset of peripheral -cells and low-grade MALT lymphoma, splenic marginal zone lymphoma and -cell chronic lymphocytic leukaemia (B-CLL). Thus, expression of adhesion molecules and chemokine receptors determines homing and dissemination of both normal and malignant -cells.

Staging of small bowel lymphoma

It is recommended that all new patients with NHLs should have their tumours classified in the first instance according to the World Health Organization (WHO) classification (Table 29.4).

Thereafter, the specific therapy in the individual patients is formulated after clinical staging with the modified Ann Arbor system taking into consideration the International Prognostic Index for the individual patient. Appropriate clinical staging includes a careful history and physical examination; CT or MRI of the chest, abdomen and pelvis; blood chemistry; full blood count; and bone marrow biopsy. The basic staging and investigation for a patient with NHL includes all these studies. Patients at high risk of CNS involvement should have a lumbar puncture for cytology of cerebrospinal fluid.

The details of the Ann Arbor staging system and the Cotswold modifications are shown in Table 29.5.

The Ann Arbor staging system does not provide reliable prognostic information for many subtypes of NHLs and is thus insufficient on its own for the formulation of the treatment regimen in the individual patient. The International Prognostic Index (IPI) is nowadays used for all types of NHLs to individualize prognosis and treatment. The reported studies with the IPI have identified factors which predict the outcome of treatment:

- age greater or less than 60 years
- Ann Arbor stages I and II vs stages III or IV
- none or one vs two or more sites of extranodal involvement by lymphoma
- Eastern Cooperative Oncology Group performance status of grade 0 or 1 vs grade 2 or greater
- a normal vs elevated lactate dehydrogenase (LDH).

The bad predictors of outcome identified by the IPI are:

- age >60 years
- performance status >2
- LDH 1 × normal
- extranodal sites >2
- stage III or IV.

As the adverse prognostic factors have the same prognostic weighting, they are added giving a score of 0–5. The individual

Table 29.4 World Health Organization classification

β-cell lymphomas	**T/NK-cell lymphomas**
Precursor β-cell	**Precursor T-cell**
B lymphoblastic	T lymphoblastic
Mature (peripheral) β-cell	**Mature (peripheral) T-cell**
Diffuse large β-cell lymphoma*	Peripheral T-cell lymphoma
Follicular lymphoma*	Not otherwise specified
Small lymphocytic lymphoma (CLL)	Angioimmunoblastic T-cell lymphoma
	Extranodal nasal NK/T-cell lymphoma
Extranodal marginal zone β-cell lymphomas of the MALT type*	Enteropathy-type T-cell lymphoma
Nodal marginal zone β-cell lymphoma	Hepatosplenic gamma-delta T-cell lymphoma
Splenic marginal zone β-cell lymphoma	Subcutaneous panniculitis-like T-cell lymphoma
Lymphoplasmacytic lymphoma	Anaplastic large T-/null-cell lymphoma
Burkitt's lymphoma	Mycosis fungoides/Sézary syndrome
	Adult T-cell lymphoma/leukaemia

*Account for more than 50% of non-Hodgkin's lymphomas.
CLL, chronic lymphocytic leukaemia; MALT, mucosa-associated lymphoid tissue; NK, natural killer.

Table 29.5 Ann Arbor staging system and the Cotswold modifications

Stage	**Features**
I	Involvement of a single lymph node region or lymphoid structure (e.g. spleen, thymus, Waldeyer's ring)
II	Involvement of two or more lymph node regions on the same side of the diaphragm
III	Involvement of lymph regions or structures on both sides of the diaphragm
IV	Involvement of extranodal site(s) beyond that designated E
For all stages	
A	No symptoms
B	Fever (>38°C), drenching sweats, weight loss (10% body weight over 6 months)
For stages I–III	
E	Involvement of a single, extranodal site contiguous or proximal to known nodal site
Cotswold modifications	
Massive mediastinal disease has been defined by the Cotswold meeting as a thoracic ratio of maximum transverse mass diameter greater than or equal to 33% of the internal transverse thoracic diameter measured at the T5/6 intervertebral disc level on chest radiography	
The number of anatomic regions involved should be indicated by a subscript (e.g. II_3)	
Stage III may be subdivided into: III_1, with or without plenic, hilar, coeliac or portal nodes; III_2, with para-aortic, iliac, mesenteric nodes	
Staging should be identified as clinical stage (CS) or pathologic stage (PS)	

patient is categorized on this score into one of four groups: (1) low risk (score of 0 or 1), (2) low to intermediate risk (score of 2), (3) high intermediate group (score of 3) and (4) high risk (score of 4 or 5). The published data on the response rate and outcome of patients demonstrate the importance of the IPI scoring (Table 29.6).

Clinical features

Small bowel lymphoma may occur at any age except in infancy, but the peak incidence is in the sixth decade. A smaller peak is encountered in the first to third decades. In general, small bowel NHLs are commoner in males, with a reported sex ratio of 2:1. The presentation may be acute or insidious. In both the Middle East and the West these diseases may present as a surgical emergency with intestinal obstruction or a perforation leading to peritonitis (6–8%). The intestinal obstruction may be due to intramural obstruction by a circumferential lesion or intussusception. The latter is particularly likely to occur with the ileocaecal tumours of childhood.

The most common presenting symptom of small bowel NHLs is pain in the abdomen (80%). Other presenting symptoms include change in bowel habits [diarrhoea or constipation (15%), palpable abdominal mass (15%) and blood in the stool (15%)]. The most common site of the disease is the terminal ileum or ileocaecum (40%) with the rest distributed at various other sites of the small intestine, with some patients (10%) having multifocal areas of involvement. Some patients (5%) give a history of Crohn's disease.

In addition to abdominal pain, the chronic manifestations include malaise, weight loss and anaemia. The anaemia may be normochromic (chronic disease) or hypochromic microcytic (chronic occult bleeding). The erythrocyte sedimentation rate (ESR) is elevated and hypoproteinaemia is frequently present and results from a protein-losing enteropathy. In patients with coeliac disease the enteropathy-associated lymphoma tends to occur in the fifth to the seventh decade. The symptoms of coeliac disease previously controlled by dietary management return and the patients complain of abdominal pain and diarrhoea with rapid weight loss. Perforation leading to peritonitis is a common presentation in these patients.

In the Middle East, IPSID is associated with growth retardation, malabsorption, bacterial overgrowth, hypoproteinaemia with oedema and ascites, and various parasitic infestations. It can also present with intestinal obstruction, perforation and massive haemorrhage. In all cases the commonest physical finding is a mobile abdominal mass.

Table 29.6 Relation between IPI score, response rate and survival

Risk	IPI score	Complete response (%)	Overall 5 year survival (%)
Low	0–1	87	73
Low/intermediate	2	67	51
Intermediate/high	3	55	43
High	>4	44	26

IPI, International Prognostic Index.

Diagnosis

The diagnosis is usually established by means of a small bowel contrast enema, although in recent years small bowel capsule and single-balloon endoscopy DBe or are being used more often. The advantage of balloon endoscopy lies in the provision of biopsy for histological characterization of the lesion. Other investigations used for establishing diagnosis include abdominal ultrasound, CT and MRI scanning and diagnostic laparoscopy. In cases presenting acutely (obstruction/perforation), the diagnosis is made at emergency laparotomy. In IPSID the abnormal α-chain is detected by immunocytochemistry of tumour sections and by immunoelectrophoresis with monospecific IgA antibody of serum and duodenal juice.

Treatment

NHLs are often treatable and frequently curable. However, the choice of the most appropriate therapy in the individual patient requires accurate diagnosis and a careful staging evaluation. New patients with NHL should have their tumours classified using the WHO classification. Patients are then assigned an anatomic stage using the Ann Arbor system and individually scored by the IPI.

Nonetheless, the treatment of gastrointestinal lymphoma is not standardized and there have been few clinical trials. All patients presenting with acute abdominal disease require surgical intervention, and whenever possible the disease should be resected. Further treatment is then administered soon after recovery from the operation. This may consist of combination chemotherapy with drug regimens that are commonly used in Hodgkin's disease (CHOP, CMOPP, etc.) or radiotherapy. Radiotherapy is used less frequently as it is no more effective than chemotherapy and carries a high morbidity (bleeding and perforation).

In uncomplicated lymphoma, surgery followed by chemotherapy or radiotherapy is used for stage I and stage II of the disease. Chemotherapy alone is used for more advanced disease, but the prognosis in these cases is extremely poor. Complete remissions have been reported in patients with IPSID whose biopsy shows plasmacytic infiltration after treatment with tetracycline or cytotoxic drugs, but patients with established lymphoma are generally treated as outlined above. There is no role for surgical treatment in patients with malignant lymphomatous polyposis because of the widespread nature of the lesion, and reliance is placed on effective combination therapy.

Combined-modality therapy (surgical resection followed by chemotherapy) in patients with intermediate- or high-grade disease has resulted in improved survival rates. Chemotherapy reduces recurrence rates and metastatic disease outside the abdominopelvic cavity. Resection remains the initial treatment as it establishes the exact diagnosis (type and grade of the lymphoma), reduces the tumour burden, relieves the symptoms and prevents potential perforation. The role of radiation therapy remains controversial because of its potential morbidity. However, several reported series have shown good tolerance to radiotherapy, although most agree that it is not indicated in patients in whom complete resection has been achieved, as these patients do not derive any benefit from additional abdominopelvic radiotherapy over postoperative chemotherapy alone. In the last 10 years significant advances have been made in the systemic treatment of NHLs, which include the use of humanized monoclonal antibodies, such as rituximab, either alone or more usually with conventional chemotherapy. In some types of lymphomas, exemplified by follicular NHL, they have drastically improved cure rates.

Specific intestinal NHLs

MALT lymphomas

Mucosa-associated lymphoid tissue (MALT) lymphomas occur in a variety of organs, including the orbit, conjunctiva, salivary glands, skin, thyroid gland, lungs, stomach (most common site) and intestine. MALT lymphomas occur at a frequency of about 1.5 per 100 000 people per year in the USA and account for about 10% of all NHLs but the incidence in Western countries varies considerably depending on the prevalence of infection by *H. pylori*, genetic, dietary and environmental factors. Thus in certain regions of Italy, the frequency of MALT lymphomas is about 13 per 100 000 people per year. *Campylobacter jejuni* infection is associated with small bowel MALT lymphomas. The other infections associated with MALT lymphomas include *Borrelia burgdorferi*, *Chlamydophila psittaci* and hepatitis C virus.

Also known as extranodal marginal zone lymphomas, these are mostly low-grade tumours (predominantly small cell) with a tendency to remain localized for long periods and to metastasize late to other sites of MALT. The marginal zone of the B-follicle represents a well-defined compartment of the -cell area with a distinct cellular composition from that of the follicle centre (follicular -cells), from which it differs in its functional role in the immune response. Some tumours have a prominent large cell component. These are high-grade tumours and carry a less favourable prognosis. Microscopically the low-grade tumours form well-defined growths with deep invasion of the bowel wall and are usually single, whereas the high-grade tumours involve extensive segments of the bowel and form large strictured lesions with a tendency to ulceration. Histologically the reactive -cell follicles of the normal lymphoid tissue of the gut are surrounded and infiltrated by neoplastic small to medium -cells with an irregular nuclear contour, which resemble centrocytes. These tumours also form a characteristic 'lymphoepithelial lesion' when the neoplastic centrocyte-like cells invade and damage the crypt epithelium.

Eradication using a triple anti-*H. pylori* regimen in gastric MALT stage IE tumours achieves a complete response in approximately 60–90% of patients. Patients with tumours that are T4 node positive or tumours with adverse cytogenetics should receive radiotherapy or surgery with or without radiotherapy (see Chapter 23). High-grade lesions or large cell tumours with a minor low-grade MALT component are treated with CHOP (cyclophosphamide, doxorubicin, vincristine and prednisone) chemotherapy. This is also used in patients in whom eradication therapy against *H. pylori f*ails. Disseminated MALT lymphomas are incurable and are treated primarily with palliative chemotherapy.

Centrocytic lymphoma (mantle zone lymphoma)

Centrocytic lymphoma is an intermediate-grade lymphoma. Originally thought to arise from small cleaved cells in the germinal centre, subsequent immunological studies have indicated an origin from mantle zone CD5+ B-lymphocytes producing IgM and/or IgD. Genuine cases of centrocytic lymphoma express the *bcl*-1 oncogene. In the recent classification, centrocytic lymphoma is placed in the category of diffuse small cleaved cell myelorid lymphoma and is often referred to as intermediately differentiated lymphocytic lymphoma or as mantle zone lymphoma. The pattern of growth is usually diffuse, often with a follicular appearance.

Mantle zone lymphomas account for 15% of primary intestinal lymphomas. They form superficial diffuse lesions which do not invade the bowel wall deeply, resulting in a convoluted appearance of the mucosa, suggesting multiple polyps, and hence the alternative descriptive clinical name of malignant lymphomatous polyposis. As opposed to other -cell tumours, surgical excision is not appropriate and the treatment is by chemotherapy. Another feature which distinguishes these primary gut tumours is the development of a centrocytic leukaemia and peripheral lymphadenopathy in most patients.

Mediterranean lymphoma

This affects mainly young adults with almost equal sex incidence and involves predominantly the proximal small intestine and is usually associated with chronic diarrhoea and abdominal pain. In 1978, the WHO recommended the term IPSID for the syndrome associated with Mediterranean lymphoma because at that time the disease, especially in its early stages, was not considered to be a truly malignant lymphoma. Many of the patients with IPSID syndrome were found to have variable levels of abnormal immunoglobulin in the serum or other body fluids, which were identified subsequently as truncated heavy chains, hence the alternative synonym of heavy-chain disease (HCD).

The diseased mucosa shows a heavy plasma cell and B-lymphocyte infiltration. The abnormal plasma cells secrete a fragment of IgA (α or heavy chain), which can be detected in the plasma and duodenal juice of affected patients. In some patients suffering from IPSID a malignant -cell lymphoma develops. The initial uncertainty regarding the nature of IPSID (benign disease which undergoes subsequent malignant change) has been resolved by subsequent molecular genetics studies which have confirmed the abnormal *Ig* gene rearrangement in both IPSID and Mediterranean lymphoma, indicating that IPSID is in fact a neoplastic condition. The WHO classification does not recognize the so-called 'non-secretory' IPSID as a specific entity. This variant is characterized by the proliferation of small centrocyte-like lymphoid cells lacking the plasmacytic differentiation observed in the secretory form of HCD. Both 'secretory' and 'non-secretory' forms are common in the same geographic area.

The aetiology remains unknown, although there is good circumstantial evidence for the involvement of various infective agents and their toxins and parasitic infestation, which account for the regional incidence of the disease. As *C. jejnui* has been isolated from patients with IPSID who respond dramatically to treatment by antibiotics, the current hypothesis is that continuous antigenic stimulation within the gut due to *C. jejuni* infection induces the proliferation of IgA-producing plasma cells in the lamina propria of the small intestine.

Patients with IPSID have an acquired immune deficiency (humoral and cellular) and usually have low serum Ig levels. This deficiency and the subsequent development of IPSID have been attributed to the modulatory effects of *Vibrio cholerae* toxins acquired during the epidemic that swept the same IPSID geographic areas during the early 1960s.

Infestations with various parasites (*Ascaris lumbricoides*, *Giardia lamblia*, *Strongyloides stercoralis*, *Trichuris trichiura*) can cause mutations in -cells with their consequent development and differentiation into aberrant plasma cells producing truncated heavy chain proteins. Continuing proliferation of these aberrant cells within the mucosa gives rise to the clinicopathological picture of HCD and eventual progression to the full-blown lymphomatous phase of IPSID. Genetic factors may be involved as evidenced by the association of IPSID with human leucocyte antigens (HLAs) AW19, A9 and B12, and the blood group B. Genetic predisposition is further substantiated by the development of IPSID in relatives living apart from patients with the disease. Occult defects of the cellular and humoral immunity have been detected in first-degree healthy relatives of patients with IPSID. However, it is difficult to be certain whether these observations indicate a genetic predisposition or are the result of shared environmental factors.

The macroscopic features of Mediterranean lymphoma are highly variable and range from diffuse thickening of the upper small intestine with enlargement of the associated mesenteric lymph nodes to localized, often multiple tumours. The histological features vary according to the stage of disease. Initially there is a diffuse infiltration of lamina propria by malignant plasma cells (stage A). Thereafter aggregations of centrocyte-like -cells (stage B) and immunoblasts (stage C) appear and the picture becomes polymorphic. Although often extensive, Mediterranean lymphoma tends to remain confined until the late stages of the disease.

Burkitt-type lymphoma

There are two types of this primary high-grade -cell lymphoma: African or endemic, and sporadic or non-African type, which occurs in Western countries. In both types, the diagnosis is made on biopsy.

The primary gene involved in 90% of patients with Burkitt's lymphoma is the c-*myc* oncogene. It usually results from a translocation of genetic material between the long arm of human chromosomes 8 and 14; less commonly the translocation involves chromosomes 2 and 22 or chromosomes 2 and 8. The c-*myc* oncogene appears to stimulate another gene, the *HMG-I/Y* gene, which then triggers the neoplastic transformation of the -cells. Burkitt's lymphoma is also associated with certain oncogenic viruses, specifically the EBV in endemic Burkitt's and HIV in the sporadic form. Currently, however, the precise role of EBV in Burkitt's lymphoma is still being investigated, because the virus is less common in patients outside Africa. In the USA, about 25% of children and 40% of adult AIDS patients with Burkitt's have the EBV.

The endemic or African type was first described in 1956, by an Irish surgeon (Dennis Burkitt) working in central Africa. The disease, which is very common in children (average age 7–8 years) who are infected with the EBV and usually suffer from chronic malaria infections, is a non-Hodgkin's -cell lymphoma. It is thought that the EBV reduces the resistance and thereby leads to the malignant transformation of the infected -cells. Endemic Burkitt's lymphoma is responsible for 50% of cancer deaths in children in Uganda and central Africa. The disease is characterized by rapid enlargement of the patient's jaw, loosening of the teeth, protruding eyeballs, or abdominal, renal or ovarian tumours.

The sporadic or non-African type of Burkitt's lymphoma which occurs in other countries is also common in children. In this form of the disease, which is more common in males, the EBV infection is acquired because of reduced immunity. These children may present with a facial tumour but much more commonly (90%) with an abdominal mass in the right lower quadrant (ileocaecal region), pain, vomiting and ascites. Involvement of the bone marrow with thrombocytopenia leads to bleeding episodes and anaemia. Other patients may develop tumours in the face/nasal sinuses, testes, ovaries and skin.

In adults, the first symptoms of Burkitt's lymphoma include lymph node enlargement, swollen and painful abdomen, itching, fever and asthenia. The lymphoma frequently produces abdominal enlargement due to hepatosplenomegaly. Burkitt's lymphoma is 1000 times more common in AIDS patients than in the general population, and approximately 2% of AIDS patients develop Burkitt's lymphoma, with the majority having advanced (stage IV) disease at the time of diagnosis.

The treatment of Burkitt's lymphoma is multimodal: surgery in selected cases and chemoradiotherapy. Bulky abdominal tumours and those causing symptoms/complications are removed surgically before systemic therapy is instituted. In endemic disease, cyclophosphamide is used (orally or intravenously). Sporadic disease in children is treated with a short course of high-dose combination chemotherapy using cyclophosphamide with methotrexate, vincristine, prednisone and doxorubicin. Radiotherapy to the CNS and intrathecal methotrexate are used as prophylaxis against the spread of the lymphoma to the CNS and spine.

Adults with sporadic Burkitt's lymphoma are treated with a combination of radiotherapy and chemotherapy. High-dose chemotherapy with CODOX-M/IVAC, (cyclophosphamide, vincristine, doxorubicin, methotrexate, ifosfamide, etoposide and cytarabine) is commonly used. However, adults with AIDS because of their poor general state are usually given low-dose chemotherapy. Newer methods of treatment include bone marrow or stem cell transplantation and monoclonal antibodies, e.g. rituximab. Early results with rituximab combined with standard chemotherapy have shown improved remission rates and survival.

The prognosis in children is generally good. Children with African Burkitt's usually obtain a good response with cyclophosphamide. In the USA, 80% of children treated for early-stage disease obtain a good response and remain free from relapse for several years after treatment. Therapy with the CODOX-M/IVAC regimen has been reported to achieve cure rates of up to 90% in both children and adults. In patients with AIDS, the factors which influence outcome include the CD4 lymphocyte count; the presence of opportunistic infections, bone marrow involvement, spread beyond the lymph nodes, age and the patient's performance status. The prognosis is bad in patients with opportunistic infections, CD4 count <200, age above 35 years and low general health/performance status. The average survival of HIV-positive patients with Burkitt's lymphoma is 6 months.

Enteropathy-associated T-cell lymphoma

This T-cell tumour is aetiologically related to coeliac disease. However, it can arise in the absence of this condition. Enteropathy-associated T-cell lymphoma (EATL) accounts for 10–25% of all primary NHLs of the small bowel. The lesion is often multifocal and can form ulcers, strictures, plaques, nodules or diffuse thickening. In general the prognosis is poor and the disease tends to become disseminated at an early stage. EATL is a T-cell lymphoma of the intraepithelial T lymphocytes, composed of large lymphoid cells and often associated with necrosis and an inflammatory background, including large numbers of histiocytes and eosinophils, hence the former name of malignant histicyosis of the small intestine. The adjacent small intestinal mucosa shows villous atrophy, crypt hyperplasia, increased lamina propria lymphocytes and plasma cells. In 10–20% of cases, the lymphoma is composed of monomorphic medium-sized cells with no inflammatory background and rare necrosis (type II EATL).

EATL is encountered with greater frequency in those areas with a high prevalence of coeliac disease (northern Europe). It has been estimated that 2–3% of patients suffering from coeliac disease will develop an intestinal lymphoma, and in the majority this has the T-immunophenotype. Coeliac disease patients have a standardized incidence ratio of 5.81 for all NHLs and 40.51 for small bowel lymphomas. The interval between diagnosis of coeliac disease and development of EATL varies considerably.

Another condition associated with EATL is ulcerative jejunitis. Small bowel NHL T-cell lymphoma is also the most frequent extranodal site of NHL developing in solid-organ transplanted recipients, especially patients with renal graft transplants.

EATL usually develops in the jejunum or ileum as one or more ulcerating mucosal lesions that invade the wall of the intestine and frequently cause perforation. The neoplastic cells are medium–large with round or indented nuclei, prominent nucleoli with a rim of pale-staining cytoplasm. Less frequently, they are pleomorphic, mimicking anaplastic large cell lymphoma. An inflammatory background is often present, consisting of histiocytes and eosinophils. The intestinal mucosa adjacent to the neoplasm frequently shows enteropathy with villous atrophy, crypt hyperplasia, increased lamina propria lymphocytes and plasma cells and intraepithelial lymphocytosis. In type II EATL, the neoplastic cells are medium-sized with darkly stained nuclei and a moderate rim of pale cytoplasm. However, this type is not accompanied by an inflammatory background and necrosis is less evident than in classical EATL.

EATL classically presents in adults, usually with a history of coeliac disease. The presentation is with abdominal pain, often associated with jejunal perforation, weight loss, diarrhoea or bowel obstruction. Thus many cases are diagnosed at laparotomy. EATL is an aggressive lymphoma that is usually fatal and its treatment is not standardized. The role of surgery is limited to debulking or resection of masses with high risk of obstruction or perforation. Radiation therapy is indicated in some patients presenting with bulky disease, rectal lymphoma or after incomplete resection. Combined treatment with primary debulking resection and systemic anthracycline-containing chemotherapy, which may or may not be followed by radiation therapy, is used in fit patients. However, relapses after chemotherapy are common (80% of patients). Additionally, many patients are unable to complete chemotherapy and do not receive radiotherapy because of rapid progression of disease, poor nutritional and performance status and local/systemic complications. Some studies have suggested a better prognosis in patients after macroscopically complete resection and the use of chemotherapy. Others have used high-dose chemotherapy supported by autologous stem cell transplantation with encouraging results, including response rates in 50% of patients.

Alemtuzumab (humanized anti-CD52 monoclonal antibody) has also been used in patients in combination with gemcitabine or CHOP chemotherapy, and initial results are promising, although further phase II studies are needed.

Primary gastrointestinal follicular lymphoma

Follicular lymphoma is the second most common subtype of NHL and accounts for 20–25% of cases in the USA and Europe. The clinical course of the disease is typically indolent, although many patients relapse after treatment. The advent of anti-CD20 monoclonal antibodies (rituximab) has profoundly changed the therapeutic management of the disease.

Gastrointestinal follicular lymphoma (GI-FL) arises from antigen-responsive -cells predominantly of the duodenum or the small intestine. The disease may be unifocal or multifocal. As with other non-Hodgkin's follicular lymphoma, it has a low tendency to disseminate outside the gastrointestinal tract. GI-FLs resemble nodal FLs with respect to morphology and expression of typical markers such as CD10, CD38 and BCL-6. The tumour cells of GI-FL express high levels of the antiapoptosis protein BCL-2 and carry mutated immunoglogulin heavy-chain genes. In contrast to nodal FLs, the GI-FLs express the $\alpha4\beta7$ integrin, a mucosa-homing receptor also expressed by normal intestinal B- and T-lymphocytes and by low-grade MALT lymphomas.

GI-FL occurs most frequently in the duodenum. The Follicular Lymphoma International Prognostic Index (FLIPI) identifies patient subgroups with predictable outcome. FLIPI is based on five simple independent risk factors: haemoglobin <12 g/dL, serum LDH higher than upper normal value, Ann Arbor stage III–IV, number of nodal sites >4 and age >60 years. The prognosis of follicular lymphoma has improved in recent years with marked increase in the 5-year overall survival (from 64% to 95%). This improvement has been predominantly observed in patients with advanced (stage IV) disease and in younger patients. It is the result of the introduction of monoclonal antibody therapy. Prior to this, patients with stage III–IV disease were treated with systemic chemotherapy including various combinations of anthracycline with alkylating agents and interferon administration (either during chemotherapy or for maintenance). The advances in the last 10 years include (1) use of the monoclonal antibody rituximab alone as first-line therapy either as a short course or with maintenance; (2) rituximab as first-line therapy in combination with chemotherapy; (3) front-line therapy with radioimmunoconjugates, alone or after chemotherapy.

The reported results of front-line therapy with rituximab have shown a response rate of 75% to a standard weekly four-dose programme, with 50% of patients obtaining a complete response. However, in the absence of further treatment, the median time to disease progression is limited to a maximum of 24 months. Improved outcome is obtained with repeated rituximab courses consisting of 4 weekly infusions every 6 months for 2 years. This increases the median progression-free survival (PFS) to 34 months. A randomized study has shown a significantly improved PFS time in the maintenance arm (31 vs 7 months), with these patients being more likely to achieve a lasting. Several randomized studies on the use of conventional chemotherapy such as CVP (cyclophosphamide, vincristine, prednisone) with or without rituximab have confirmed improvement in response rates, PFS and overall survival favouring the rituximab-containing arm. In another trial (ECOG 4496), patients who did not progress after six to eight courses of the CVP regimen were randomly allocated to no further treatment vs consolidation with four cycles of 4 weekly rituximab infusions administered every 6 months for 2 years. This study demonstrated a significant improvement of 3 year PFS in the rituximab consolidation arm. Altogether, these studies demonstrate that a first-line treatment with rituximab with or after chemotherapy improves the outcome of these patients. The use of radioimmunoconjugates in the first-line treatment of patients with follicular lymphoma, either as a single agent or as consolidation therapy, has been demonstrated to improve patient outcome. Thus front-line therapy with ^{131}I-tositumomab achieves a very high response rate (95%), with 75% of patients achieving a complete response, a 5-year PFS of 59% and limited toxicity.

Inflammatory conditions of the small bowel

Crohn's disease

Crohn's disease is an idiopathic chronic inflammatory condition which can affect any part of the gastrointestinal tract and may also be associated with systemic manifestations. The disease is localized in the ileocolic region in 60% of patients, to the small bowel alone in 20% and to the colon alone in a further 20%. Perianal disease is common and may accompany more proximal disease. Colorectal and perianal disease are dealt with in chapter 30. Cases of Crohn's disease of the mouth, oesophagus and stomach are extremely rare. The disease is most common in North America and northern Europe, and although the prevalence is increasing in southern Europe it is relatively uncommon in other areas of the world. In the Far East it is almost never encountered.

Pathology

Irrespective of the site of involvement, Crohn's disease is a segmental condition with areas of involvement that are sharply demarcated from the contiguous normal bowel, at least to naked eye appearances. Particularly in small bowel disease, there may be several diseased segments with normal intervening bowel (skip lesions) but the number of such lesions is highly variable.

Macroscopic appearance

In early stages, the disease appears as mucosal inflammation with small aphthoid ulcers. In more advanced disease, the serosal surface of the affected bowel is granular and inflamed with a tendency for it to be encroached by mesenteric fat, so that the intestine may at times be buried within a swollen oedematous and foreshortened mesentery. On palpation, the involved areas feel heavy, thickened and firm as a result of the transmural inflammation that is usually associated with narrowing of the bowel lumen. Close inspection of the opened bowel reveals separation of the usual anatomical layers by fibrosis, which is particularly marked in the submucosal and subserosal layers (Figure 29.18). The mucosal oedema accounts for the characteristic cobblestone appearance of the mucosa (Figure 29.19). This oedema is followed by sloughing and linear ulceration of the mucosa, particularly at the mesenteric attachments (Figure 29.20). Pseudopolyps and mucosal bridges may form and the ulcers, which are typically deep and fissuring, penetrate into the muscle layers, and account for the tendency to localized perforation, adhesions and fistula formation. Regional lymphadenopathy is invariably present and usually the result of reactive hyperplasia.

Fistula formation is an important feature of Crohn's disease and accounts for substantial morbidity, long periods of debility and a 5–10% mortality following surgical treatment. The fistulous tracts may be simple or complex and often incorporate

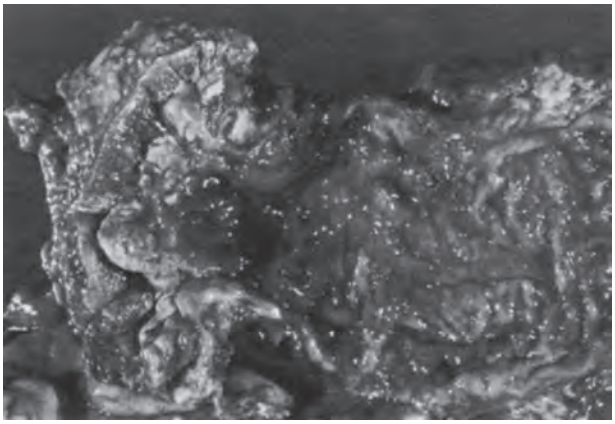


Figure 29.18 Crohn's disease of the ileocolic region. The affected bowel has been cut longitudinally to demonstrate the transmural fibrosis which is especially marked in the subserous and submucosal layers.

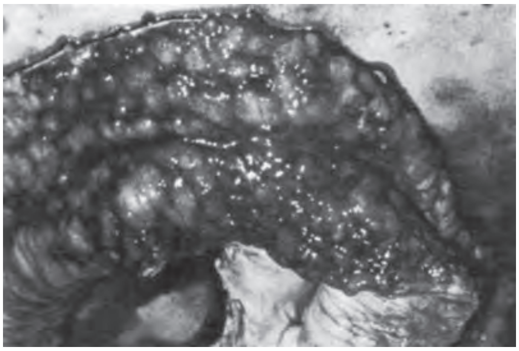


Figure 29.19 Diffuse mucosal/submucosal oedema resulting in the characteristic cobblestone appearance of the mucosa of Crohn's disease.

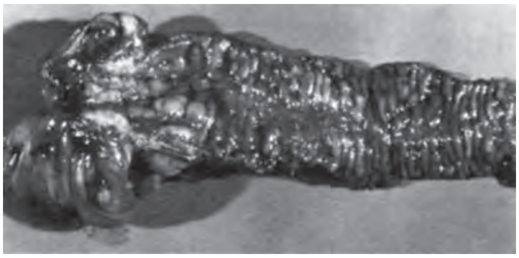


Figure 29.20 Extensive ulceration with stricture formation in Crohn's disease of the distal ileum. The longitudinal ulceration is more prominent along the mesenteric attachment.

intervening or associated abscesses. The important varieties of fistula encountered in Crohn's disease are listed below.

- *Enterocutaneous*: some 20–30% of these may heal with drainage of the associated abscess and parenteral nutrition.
- *Perianal fistulas*: these may be associated with severe anorectal disease but not in all cases.
- *Ileocolonic fistulas*: these can lead to diarrhoea and severe bacterial overgrowth of the small intestine leading to malabsorption and malnutrition.
- *Enterovesical fistulas*: these may involve the ileum or the colon or both and usually develop in patients with longstanding disease. Enterovesical fistulas are less common than the other varieties and have a reported incidence of about 5%. They usually follow a benign course and present with dysuria, pneumaturia and recurrent pyelonephritis. They do not respond to conservative measures and require surgical intervention.

Microscopic appearances

The histological features of Crohn's disease are characterized by a transmural inflammation consisting of chronic inflammatory cell infiltrates, including eosinophils, crypt abscess formation, oedema, non-caseating epitheloid cell granulomas containing giant cells (Langerhans cells), dilatation and sclerosis of the intestinal lymphatics and lymphoid aggregates with or without germinal centres. These are all found in the various layers of the bowel wall associated with epithelial regenerative changes and angiitis. These granulomas, which resemble those found in sarcoidosis, are found in the bowel wall in 50–60% of cases and to a lesser extent (25%) in the regional lymph nodes. The epithelial regenerative changes include the development of pseudopyloric gland metaplasia. The non-caseating granulomas are the most important criterion for the histological diagnosis of Crohn's disease and in their absence it can be very difficult to make the distinction between Crohn's disease and ulcerative colitis when the disease affects the large intestine.

Aetiology

The aetiology of Crohn's disease remains obscure. However, the various hypotheses may be considered under the following headings: genetic factors, environmental factors, infective agents, immunological factors and vasculitis.

Genetic factors

There is a familial tendency to develop Crohn's disease and about 10% of patients have an affected first-degree relative. There is no consistent HLA association but patients with ankylosing spondylitis are at greater risk of Crohn's disease than the general population and this association appears to be linked with tissue type HLA B27. A susceptibility locus for Crohn's disease has been mapped on chromosome 16. The gene (*NOD29*) determines the body response to some bacterial proteins. Mutations of *NOD29* are present in individuals who develop Crohn's disease. Other genes, e.g. the autophagy related 16-like 1 gene (*ATG 16L1*) and *IRGM*, which cause defective macrophage function, have also been implicated. The intestines of patients with Crohn's disease harbour higher titres of certain strains of *Escherichia coli* which have also been implicated in the pathogenesis through defective clearance of these bacteria consequent on impaired intestinal mucosal macrophage activity.

Environmental factors

The possible role of diet has been widely investigated and is thought to account for the large geographical differences in the prevalence of Crohn's disease. There are no reliable data linking fibre consumption with Crohn's disease, but several studies show a high intake of refined carbohydrates by Crohn's patients compared with appropriate controls. Smoking is probably an important factor, with smokers having a greater risk than those who have never smoked. This is in contrast with ulcerative colitis, where smoking appears to be protective. There also appears to be an increased risk of Crohn's disease in women taking oral contraceptives.

Infective agents

There is little epidemiological support for the suggestion that Crohn's disease may be infective. Investigations for evidence of clustering in time or space have not been able to demonstrate any evidence of person-to-person transmission, and studies on healthcare professionals dealing with Crohn's patients do not indicate an increased risk. However, some work suggests that low-virulence bacterial yeast or *Chlamydia* organisms may be responsible and two particular types of bacteria have been extensively investigated.

The variant or L-form bacteria have deficient cell walls and can develop from several bacterial species after exposure to antibiotics. The cell wall protein deficiency is accompanied by an alteration in virulence, antigenicity and pathogenicity. However, these variant bacteria, which can pass through filters holding back normal bacteria, can in time and under certain conditions revert back to their original form. *Pseudomonas maltophilia* is a variant bacterial species that has been isolated from intestinal Crohn's disease.

Mycobacterium paratuberculosis is an atypical mycobacterium which has been isolated from patients with Crohn's disease and on inoculation has been shown to produce granulomas in animals. In addition, patients with Crohn's disease have been shown to have IgG antibodies against this organism. However, the antibody levels do not bear any relation to disease activity and their level is not altered by resection of diseased bowel. Recent studies of DNA hybridization have revealed mycobacterial genomes in the tissues of patients with Crohn's disease, providing further indirect evidence of the possible role of the organism in Crohn's disease.

Immunological factors

Although humoral immune responses appear to be normal in patients with Crohn's disease, cell-mediated immunity is not. Some patients have impaired skin hypersensitivity, a reduction in circulating T-lymphocytes and a poor response to non-specific mitogens. It is quite possible however that these defects result from the disease itself or associated malnutrition rather than being aetiological factors. An abnormal permeability of the mucosal epithelium is well documented and it is possible that this could lead to exposure of the subepithelial immune cells to foreign protein, thus setting up an immune reaction in the wall of the intestine. This theory is supported by the favourable response of Crohn's disease to an elemental diet. Abnormal activation of the GUT immune system in Crohn's may be responsible for the chronic inflammation and ulceration. The susceptibility to abnormal activation of the immune system is genetically inherited.

Vasculitis

This hypothesis has been suggested following the demonstration of intestinal vascular pathology (vasculitis) in the evolution of Crohn's disease. In essence it implicates multifocal gastrointestinal microinfarction as the cause of Crohn's disease. This theory however remains unsubstantiated.

Clinical features

The peak incidence of Crohn's disease is in the third decade and it occurs with the same frequency in females as in males. It may however also affect children and elderly people. In the latter age group, it is frequently colonic and accompanied by diverticular disease. The clinical manifestations of Crohn's disease are extremely varied and depend on the location and the extent of disease. There are also a number of complications which can determine the presenting features. The 'syndromes' by which Crohn's disease can present include the pseudoappendicitis syndrome, small bowel obstruction, abscess formation, fistula formation, diarrhoea, growth retardation and portal venous gas.

Pseudo appendicitis syndrome

Some patients with Crohn's disease of the terminal ileum develop acute abdominal pain, the severity and location of which all simulate acute appendicitis. Some 14% of children and young adults with Crohn's disease present in this manner. This clinical presentation is of particular importance as acute terminal ileitis often due to *Yersinia* infection is often encountered in many instances and has a very different natural history from that of Crohn's disease. Approximately one in eight cases of acute terminal ileitis are due to Crohn's disease. Accordingly, in the emergency situation the treatment should be conservative. The appendix should be removed if the caecum appears normal and swabs taken from the luminal contents for bacteriological culture. Appendicectomy will prevent further diagnostic dilemmas at a later stage. However, if the caecum and the base of the appendix appear to be affected by disease, appendicectomy should not be carried out as this may lead to fistula formation.

Small bowel obstruction

Crohn's disease of the small bowel gives rise to abdominal pain because of acute or subacute small bowel obstruction (Figure 29.21). Complete small bowel obstruction is relatively rare. Patients with intermittent self-limiting obstructive episodes invariably have gross bacterial overgrowth which may give rise to further malabsorption, hypoproteinaemia and, malnutrition.

Abscess formation

This is a common presentation. Abscesses may result either from bowel perforation which may occur at the site of a deep fissuring ulcer or proximal to a stricture. It can also arise in a mass of inflamed regional lymph nodes without perforation of the bowel. Apart from the local signs and symptoms, abscess formation leads to malaise, weight loss, fever and anorexia. The commonest site of abscess formation is in the right iliac fossa. This may track into the pelvis along the psoas muscle underneath the inguinal ligament and present as a tender groin mass (psoas abscess). Free perforation of Crohn's disease into the peritoneal cavity with widespread peritonitis is extremely unusual.

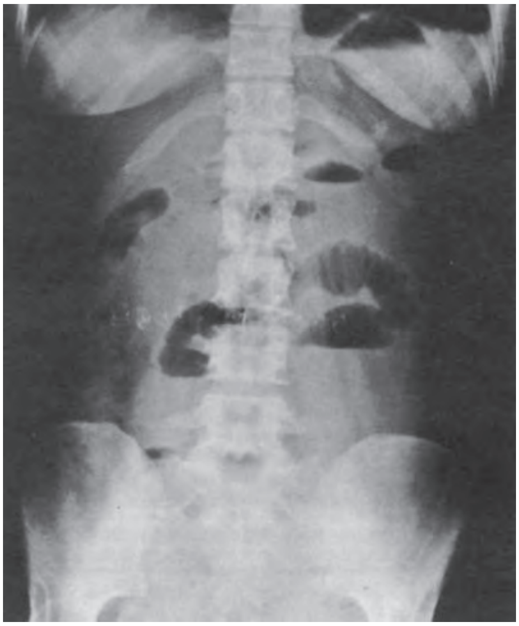


Figure 29.21 Plain erect abdominal film showing air/fluid levels in a stepladder pattern due to mechanical small bowel obstruction. The patient had had a previous right hemicolectomy for Crohn's disease. Such an obstructive picture may be due to adhesions or recurrence which is most commonly situated at the level of the previous resection.

Fistula formation

The most common type is perianal, which usually occurs in association with perianal abscesses and sinuses (Figure 29.22). These perianal complications are a significant feature of Crohn's disease and may be present months or years before intestinal symptoms are noticed. Enterocutaneous fistulas usually become evident following drainage of abdominal abscesses and can be classified as high or low output in terms of the amount of intestinal contents which discharge everyday. Spontaneous closure with parenteral nutrition is more likely in the low-output variety. Enterocolic fistulas are associated with bacterial overgrowth, malabsorption and frequently diarrhoea (Figure 29.23). Enterovesical fistulas due to Crohn's disease do not usually cause severe systemic disturbance and present with urinary symptoms (see above).

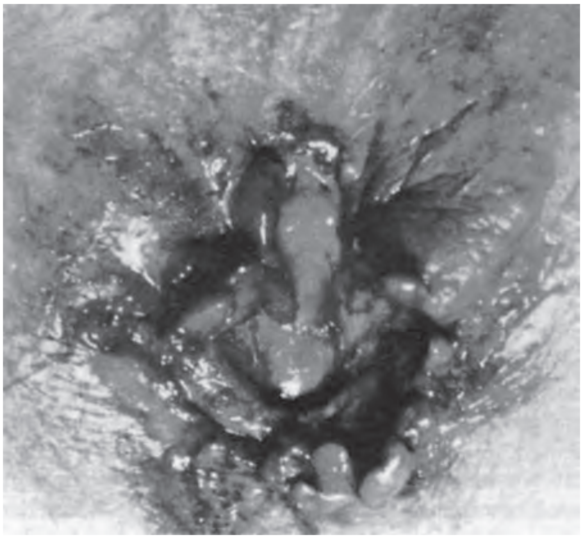


Figure 29.22 Extensive anal and perianal disease in a patient with Crohn's colitis. The patient had a high fistula *in ano*.

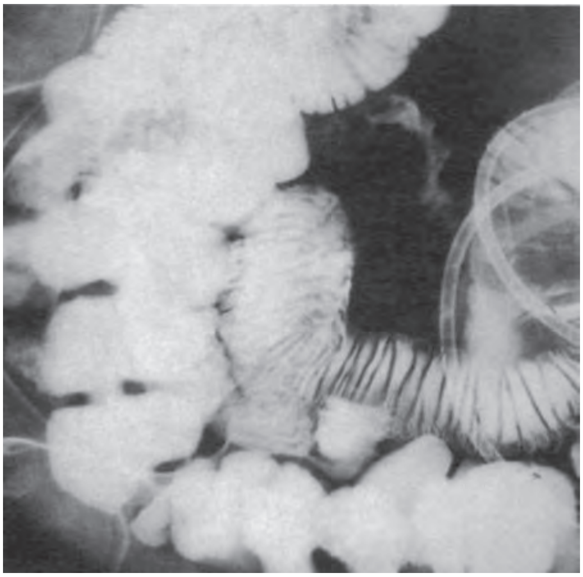


Figure 29.23 Barium study demonstrating a fistula between the small bowel and the colon in a patient with Crohn's disease.

Diarrhoea

Diarrhoea occurs in about 70–80% of patients with Crohn's disease. It is particularly common in patients with terminal ileal involvement, and this is thought to be related to the failure of the terminal ileum to reabsorb bile salts. Diarrhoea associated with colonic disease is accompanied by the passage of mucus and blood from the rectum and is very similar to the symptomatology of ulcerative colitis. Such diarrhoea is often accompanied by a protein-losing enteropathy and steatorrhoea due to bacterial overgrowth or bile salt malabsorption.

Growth retardation

This occurs in approximately 20% of all children with Crohn's disease. Failure to thrive in children with Crohn's disease has four basic components: impaired linear growth, weight loss or failure to gain weight, delayed sexual maturation and closure of epiphyses and normal endocrinological parameters. Thus thyroid function in these children is consistently normal and so is growth hormone secretion when several provocative tests are employed. The major factor in growth retardation is malnutrition which is due to failure to eat because of high-grade partial intermittent small bowel obstruction, protein and blood loss from the ulcerated inflamed areas of the small bowel and malabsorption due to bacterial overgrowth. When assessing the potential growth of these children it is crucial to evaluate their bone age and to note the stature of their parents. In this way it is possible to predict which children would be capable of responding to therapeutic intervention (including parenteral nutrition). Following an appropriate bowel resection, growth resumes the normal pattern but compensatory growth does not occur. It is therefore essential to treat these children while their bones are still capable of growth. Fortunately, in many of these teenagers closure of the epiphyses is delayed because of the associated malnutrition.

Portal venous gas

This potentially fatal complication has been described in patients with Crohn's disease. The patient develops severe toxicity with high fever and rigors and severe diarrhoea. The gas in the portal vein is best detected by abdominal CT scanning and indicates portal bacteraemia. All patients who develop this complication have active disease. Treatment of this complication is not standardized as it is a very rare complication, but broad-spectrum antibiotics active against both Gram-negative aerobes and anaerobes, steroids and intravenous fluids are generally recommended in the first instance with laparotomy and resection of the diseased segment of bowel if there is no improvement.

Diagnosis

Suspicion of Crohn's disease is based on symptoms and signs (fever, abdominal pain and tenderness, diarrhoea with or without bleeding, and anal disease). Initial laboratory screening shows elevated white blood cell counts and ESR and hypochromic microcytic anaemia. Firm diagnosis then rests on correlation of the clinical manifestations with endoscopic, radiological and histological findings.

Certain laboratory tests are important to establish the nutritional state of the patients and as an index of disease activity. A full blood count, ESR, serum electrolytes, serum proteins and in particular serum albumin are important. Decreased serum albumin, hypochromic microcytic anaemia and low serum iron may be found in as many as 50% of untreated patients. Occasionally, a megaloblastic anaemia may be observed as a result of folic acid deficiency or impaired vitamin B_{12} absorption. Hypocalcaemia, hypomagnesaemia, low serum level of zinc and low serum vitamin A may also be present reflecting impaired absorption or decreased ingestion. Frank malabsorption with steatorrhoea may be present in patients with extensive small bowel disease or bacterial overgrowth. In these patients the Schilling test or the SeHCAT test may prove useful in defining the presence of malabsorption of vitamin B_{12} and bile salts respectively. Bacterial overgrowth is documented by the hydrogen breath tests and the oral ^{14}C-glycocholate test. Important indices of disease activity are ESR, serum α-1-glycoprotein level, C reactive protein and OKT9 lymphocyte positivity.

Radiological appearances

Barium contrast studies (follow-through or small bowel enema) usually identify the disease and define its distribution. Other investigations which must be performed in suspect cases include proctoscopy and colonoscopy. Biopsy material is essential for establishing the diagnosis. CT and MRI enterography provide excellent evaluation of the severity and extent of the Crohn's disease.

In small bowel disease affected segments are most commonly demonstrated by means of a small bowel enema or follow-through. Narrowing of the lumen, nodularity of the mucosal pattern, thickening of the ileocaecal valve, mucosal irregularity and deep ulcerations perpendicular to the intestinal lumen (rose thorn ulceration) and fistula formation are all characteristic's features of Crohn's disease. Long narrow strictures result in the well-known string sign of Kantor (Figure 29.24). Skip lesions are characteristic with normal bowel in between the strictured or diseased areas. Fistulas and sinuses may be evaluated by injection of contrast medium following insertion of catheters of the appropriate size to ensure a close fit. MRI provides detailed information of the pathological anatomy of the fistula and is especially useful in complex types with multiple and branched tracts (Figure 29.25).

More recently, CE has been used for the diagnosis of Crohn's disease and is reported to provide superior results to contrast studies and CT although it carries a risk of capsule impaction in patients with strictures. Thus it is contraindicated in patients with colicky abdominal pain and other obstructive symptoms.

Complications

Although intestinal obstruction abscess formation and fistula formation are technically complications of Crohn's disease, they are so common that they are usually regarded as integral clinical features of the disease. Crohn's disease may however be accompanied by systemic manifestations of inflammatory bowel disease. These include sclerosing cholangitis, skin problems such as pyoderma gangrenosum and erythema nodosum, arthritis and uveitis.

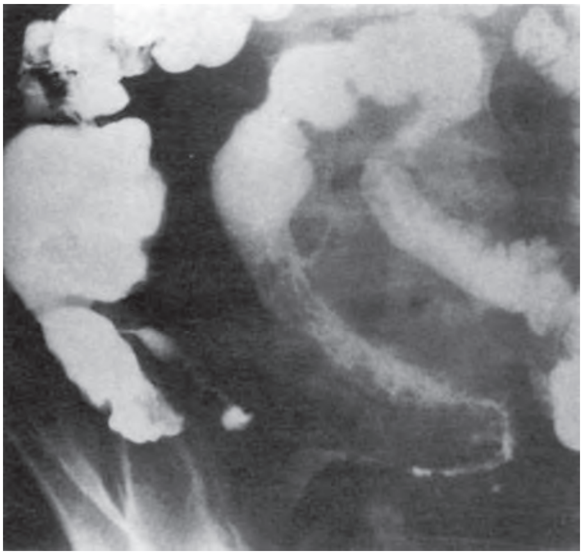


Figure 29.24 Small bowel enema showing several features of Crohn's disease of the distal ileum: stenosis (Kantor's string sign), cobblestoning of the mucosa, deep glandular ulcer (pseudodiverticulum), deep penetrating fissure ulcers, wide separation of the intestinal loops due to thickening of the bowel walls and thickening of the ileocaecal valve due to lymphatic blockage and oedema. The patient presented with intermittent cramp-like abdominal pains after meals.

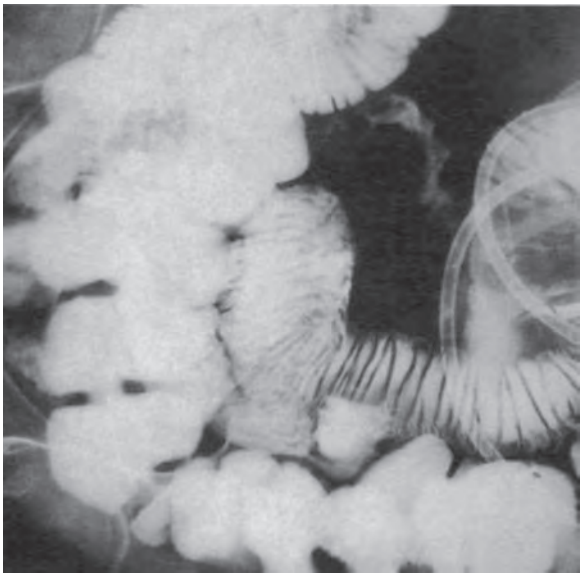


Figure 29.25 Small bowel study demonstrating a fistula between the small bowel and the colon in a patient with Crohn's disease.

In addition to enterovesical fistulas the urological complications of Crohn's disease include an obstructive uropathy and an increased incidence of renal calculi. The obstructive uropathy is caused by retroperitonitis and fibrosis leading to obstruction of the ureter (usually on the right) and is more frequent than is clinically appreciated (Figure 29.26). Patients with obstructive uropathy are usually totally asymptomatic although occasionally pyuria, bacteriuria and pain in the renal

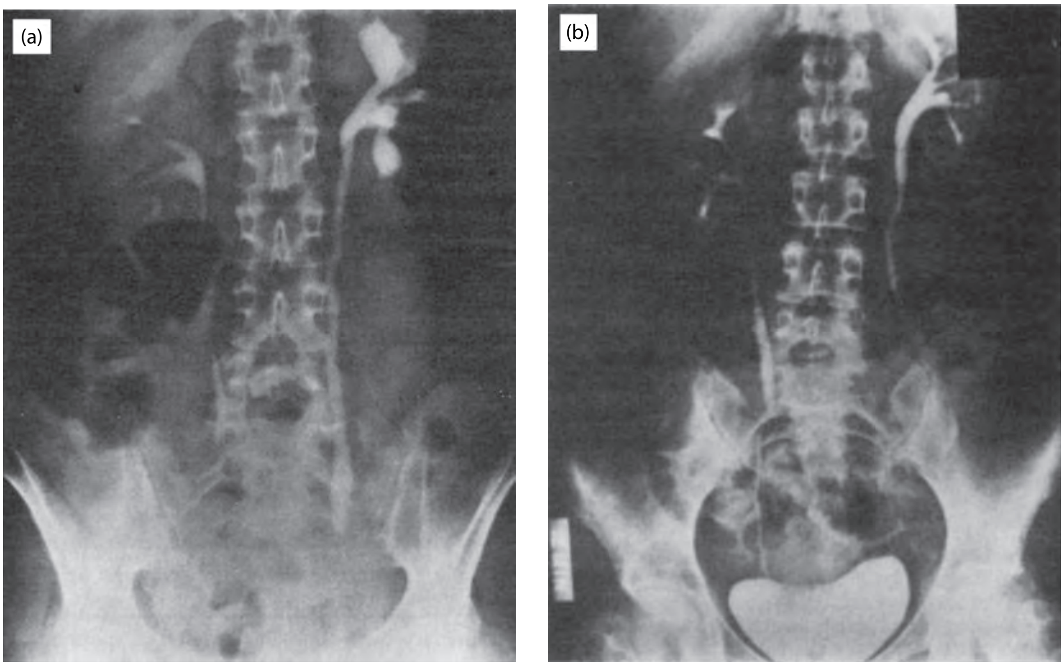


Figure 29.26 (a) Intravenous urogram demonstrating an obstructive uropathy on the left side. The patient had Crohn's disease of the ileum with fistula formation involving the sigmoid colon and resulting severe inflammation of the retroperitoneal tissue on the left side; (b) the same patient 1 month following bowel resection and ureterolysis – there is complete resolution of the left hydronephrosis and hydroureter.

angles are observed. If untreated, obstructive uropathy can lead to severe hydronephrosis, recurrent pyonephrosis and loss of functional renal tissue. Imaging of the kidneys is therefore an important part of the work-up of patients with Crohn's disease using either intravenous urography or cross-sectional imaging (ultrasound or CT scanning). Renal calculi have been reported in 10% of patients with Crohn's disease and in most of these patients the duration of disease has been longstanding. Calcium oxalate or phosphate stones account for 72% of cases and uric acid stones for 18%. Some 10% of the renal calculi cannot be classified. Only 36% of patients with renal stones have a predisposing urinary tract abnormality such as sponge kidney, obstructive uropathy or a metabolic disorder such as hypercalcaemia. Increased absorption of oxalate and increased cell turnover in the gut and concentrated urine appear to be important factors in the formation of renal calculi in patients with Crohn's disease. The prevalence of renal stones in children with Crohn's disease is said to be as high as 25%.

There is also an increased incidence of gall stone formation in patients with severe ileal Crohn's disease or following ileal resections. This is attributable to malabsorption of bile salts with interruption of the enterohepatic circulation and a consequent reduction of the bile salt pool. In addition, there is now a well-established association between Crohn's disease and small bowel adenocarcinoma. These tumours occur at a younger age and are located predominantly in the distal ileum. Cases of small bowel lymphoma and carcinoid tumours have also been reported in patients with Crohn's disease.

Treatment

There is no curative medical or surgical treatment for Crohn's disease. The aim of medical therapy is to (1) induce remission, (2) prevent relapse, (3) minimize side effects of medication and (4) improve the quality of life. Medical treatment can be further subdivided into dietary modifications, parenteral nutrition and drug treatment.

Selection of treatment regimen has to be individualized and depends on disease severity, disease location and presence of complications. There are two contrasting approaches to the medical therapy in Crohn's disease: the traditional step-up approach and the more recent top-down management. In the former regimen, initial treatment is with the least toxic agents for mild disease, with more aggressive therapy being reserved for severe disease, or for patients who fail to respond to the initial medication. The more recent top-down approach involves early aggressive therapy from the start aimed to decrease exposure to anti-inflammatory agents and use of agents which enhance mucosal healing, thereby preventing future complications. This approach has been gaining favour in recent years.

Dietary modification

In patients with extensive disease the first priority is to correct deficiencies including vitamins (especially A), magnesium and zinc. Iron-deficiency anaemia is treated with oral supplements preferably in the glutamate form. There are various reports of benefit derived from low-fibre, low-refined-sugar diets and from the use of exclusion diets. In the latter, patients are started on distilled water and then items of food introduced and maintained or withdrawn according to whether or not they aggravate the condition. It appears that the most common dietary components which require exclusion are dairy products and wheat. There is some evidence for the benefit of an elemental diet (e.g. Vivonex, Flexical or Ensure) in suppressing mucosal inflammation. The effect of elemental diets is thought to result from a reduction of the foreign protein load presented to the inflamed mucosa. The problems with long-term use of elemental diets include their expense and their unpalatability, which reduces compliance in the long term. Dietary supplements of calcium, folate and vitamin B_{12} are needed in patients with malabsorption. The use of antimotility agents (diphenoxylate and atropine, loperamide) and antispasmotics may relieve abdominal colicky attacks and diarrhoea.

Parenteral nutrition

This is of value in the following situations:

- intestinal failure with gross nutritional deficiencies and severe hypoalbuminaemia
- malnourished patients prior to surgery
- in obtaining remission in patients with severe active disease
- in patients with enterocutaneous fistulas.

Long-term home parenteral nutrition is reserved for patients with extensive small bowel disease or resection and intestinal failure. The patients are taught to self-administer the parenteral feeds and to look after their central lines and it offers a reasonable quality of life. Some patients require only intermittent supplemental intravenous feeding and manage reasonably well with modified or elemental diets in the intervening periods. Home parenteral nutrition is supervised by the medical and nursing staff of specialized nutrition units.

Intravenous feeding is necessary on a short-term basis for patients with complications requiring elective surgical intervention but who are malnourished. Improvement to the nutritional status and elevation of the serum albumin above 3 g/L is essential for the safe healing of intestinal anastomoses. Some 30% of enterocutaneous fistulas due to Crohn's disease may heal spontaneously with the use of parenteral nutrition, but persistence of a fistula beyond 6 weeks indicates failure of this approach and the need for surgical intervention.

Some advocate cessation of oral feeding and total parenteral nutrition for several weeks to achieve remission in active Crohn's disease. Although the immediate results of this approach are good, relapse is almost invariable within 3 months of resumption of oral intake.

Drug therapy

There is a wide range of drugs which are used in the management of Crohn's disease, indicating that no one drug is universally effective and that the treatment has to be individualized in accordance with the severity of the disease of the patient. All the medications fall in to five classes: (1) anti-inflammatory agents, (2) predominantly immunosuppressive drugs, (3) anti-tumour necrosis factor (TNF)-α agents, (4) anti-α-4 integrin and (5) methotrexate, which is both an immune-depressant and an anti-inflammatory drug.

Anti-inflammatory drugs

These agents decrease intestinal inflammation and include (1) 5-aminosalicylic acid (5-ASA) compounds such as sulfasalazine and mesalamine, which act by direct contact (as topical agents) with the inflamed mucosa, (2) corticosteroids as systemic agents to decrease inflammatory response by the body and (3) antibiotics. Systemic steroids incur significant side effects with long-term use. However, budesonide (semisynthetic topical corticosteroid), which acts on direct contact with the inflamed mucosa, has fewer side effects. The antibiotics used include metronidazole (Flagyl) and ciprofloxacin. Both decrease the inflammation in Crohn's disease although exact their mechanism remains unknown.

5-Aminosalicylic acid drugs

Oral 5-ASA is effective in both Crohn's disease and ulcerativcolitis. Mesalamine enema contains 5-ASA and is used in rectal disease, but, for proximal and ileal disease, 5-ASA must be taken orally. When administered orally, most of the 5-ASA is absorbed in the stomach and the upper small intestine, and very little 5-ASA reaches the ileum and colon. Hence, modified 5-ASA, which is not absorbed by the stomach and the upper small bowel, is used, e.g. sulfasalazine (5-ASA linked to sulfapyridine). Sulfasalazine is effective in inducing remissions in patients with mild to moderate disease. It is also used to prolong remissions. When sulfasalazine reaches the ileum and the colon, intestinal bacteria release the active 5-ASA. Most of the side effects of sulfasalazine are due to the sulfapyridine moiety and include nausea, heartburn, headache, anaemia, skin rashes, and, rarely, hepatitis and nephropathy. In males, sulfasalazine can reduce the sperm count. The newer 5-ASA compounds, e.g. mesalamine, do not have the sulfapyridine moiety and thus have fewer side effects than sulfasalazine, and, for this reason, are nowadays used more frequently. Asacol is an oral preparation consisting of 5-ASA coated with acrylic resin which prevents absorption of the 5-ASA in the stomach and upper small intestine. The resin dissolves as it reaches the terminal ileum and colon, releasing the active 5-ASA. Asacol is used for inducing remission primarily in patients with mild to moderate ulcerative colitis but several reports have also documented its efficacy in Crohn's ileitis and ileocolitis, and in maintaining remission from this disease. Pentasa is a capsule containing only 5-ASA. The 5-ASA is released slowly as the capsule travels down the small intestine and is active in terminal ileal Crohn's disease and the colon. It is currently the most commonly used 5-ASA compound for mild to moderate small bowel Crohn's disease. Pentasa is also used to maintain remission and reduce recurrence in patients after surgical excision of Crohn's disease. Olsalazine is a special formulation of bound 5-ASA molecules which prevents its absorption in the stomach and upper small intestine. The intestinal bacteria then break the chemical bonds, releasing the active 5-ASA molecules in the terminal ileum and colon where the bacterial counts are high. Olsalazine is effective for maintenance of remission in ulcerative colitis but it induces diarrhoea and, for this reason, it is not used often nowadays. Balsalazide capsules contain 5-ASA linked to an inert molecule which prevents absorption of the 5-ASA until it reaches the terminal ileum and colon, where bacterial activity splits the compound, releasing the 5-ASA. 5-ASA enemas and suppositories are used in the treatment of ulcerative colitis involving only the distal colon. They are also effective in treating rectal Crohn's disease. 5-ASA in suppositories is used for the same indications and may be useful in maintaining remission in patients with distal ulcerative colitis.

Corticosteroids

Steroids are used in patients with moderate to severe Crohn's disease and ulcerative colitis. Administered intravenously and orally, they remain the most appropriate treatment for severe disease or cases not responding to 5-ASA therapy. In some patients, they may also be necessary to maintain remission. Corticosteroids are indicated in Crohn's disease, ulcerative colitis and Crohn's colitis. In critically ill patients, corticosteroids are administered by intravenous injection. In patients with proctitis, hydrocortisone enemas are use to control the rectal inflammation. Although corticosteroids exert a much faster action than 5-ASA in inducing a remission, they are much less useful than 5-ASA in preventing relapse of the disease in both Crohn's disease and ulcerative colitis. The problem with corticosteroid therapy relates to their adverse side effects which are frequent and often severe especially during prolonged high-dose therapy. Thus, whenever possible, short courses of corticosteroids are used. The important side effects of corticosteroid therapy include fluid retention with the development of a moon face, acne, hirsutism, diabetes, hypertension, cataracts, glaucoma, immunosuppression with increased risk of infection, muscle weakness, depression, insomnia, personality change, osteoporosis and aseptic necrosis of the hip and knee joints (in 3–4% of patients). Another potentially fatal complication of long-term steroid therapy is adrenal atrophy such that acute adrenal insufficiency occurs on withdrawal of treatment. Thus corticosteroids should always be reduced gradually to enable recovery of adrenal function. Even after safe withdrawal of corticosteroids, the ability of the adrenal glands to produce cortisol can remain depressed for several months up to 2 years. This hypocorticalism may not be able to produce increased amounts of cortisol in conditions of stress (accidents, surgery, and infections) and thus results in adrenal insufficiency.

The majority of patients with Crohn's disease respond to an oral dose of 40–60 mg of prednisone with an improvement in symptoms within a few weeks. Once symptomatic improvement has been achieved, the dose is reduced by 5–10 mg per week to 20 mg daily. The dose is thereafter reduced at a slower rate until cessation of treatment. Aside from obviating the risks of acute adrenal insufficiency, this gradual reduction also reduces the risk of relapse of the inflammation. In patients who achieve remission with corticosteroids, 5-ASA compounds often are continued alone to maintain remission.

In patients whose symptoms return during gradual withdrawal of prednisone, the dose is increased slightly until control of symptoms is achieved, when reduction of corticosteroids can be resumed gradually. Unfortunately, many patients become corticosteroid dependent. In such patients and in patients who are unresponsive to corticosteroids and other anti-inflammatory medications, immunomodulator medications or surgery should be considered.

Budesonide (semisynthetic steroid) is a potent anti-inflammatory drug with a topical action on the inflamed tissues in Crohn's disease and has no systemic effect. On absorption, it is rapidly metabolized in the liver (first pass) to inactive compounds. Budesonide capsules contain granules (Entocort) that enable slow release of the drug into the ileum and the colon. In a double-blind study involving patients with Crohn's ileitis and/or Crohn's disease of the right colon, budesonide was found to be more effective than Pentasa in inducing remissions. Other studies have shown however that although it is no more effective than corticosteroids in inducing remission in Crohn's disease, it has fewer side effects. Trials have indicated that budesonide 5 mg per day as a single morning dose is as effective as 40 mg of prednisolone daily, and because it causes fewer side effects it is commonly used in the management of patients with active Crohn's disease. Budesonide also suppresses the adrenal glands but to a lesser extent than systemic corticosteroids. Budesonide is not effective in maintaining remission in patients with Crohn's disease.

Antibiotics

These include metronidazole (Flagyl) and ciprofloxacin. Metronidazole has some activity in the treatment of Crohn's colitis and is particularly useful in patients with perianal fistulas. Both antibiotics can induce symptomatic remission in active Crohn's disease. However, chronic use of metronidazole in doses higher than 1 gram daily can lead to a peripheral neuropathy. Patients on metronidazole must avoid alcohol because this combination causes severe nausea, vomiting, cramps, flushing and headache. Ciprofloxacin is used in the treatment of Crohn's disease either alone or in combination with metronidazole.

Symptomatic treatment for diarrhoea includes the administration of loperamide or codeine phosphate and non-addictive analgesics for the abdominal pain. Cholestyramine may improve steatorrhoea in patients with extensive ileal disease or resection.

Immunosuppressive drugs

These are used to abrogate the abnormally activated immune response in severe Crohn's disease. Their risk is an increased susceptibility to infections. The drugs in this category used in Crohn's disease are: 6-mercaptopurine (6-MP), azathioprine, methotrexate and anti-TNF-α antibodies, e.g. infliximab, adalimumab, certolizumab and natalizumab.

Azathioprine (Imuran) and 6-MP have been used for many years in the treatment of patients with moderate to severe Crohn's disease and ulcerative colitis. Azathioprine is chemically related to 6-MP and is converted into 6-MP in the body. They are indicated in:

- severe Crohn's disease and ulcerative colitis not responding to steroids
- in patients who develop serious corticosteroid-related side effects
- in patients who develop corticosteroid dependency (patients become unable to discontinue steroid therapy without relapse of their disease)
- to maintain remission.

In patients with severe Crohn's disease characterized with frequent relapses, despite 5-ASA, low-dose azathioprine and 6-MP are used to maintain remission. Perianal fistulas usually respond to treatment with metronidazole, but this treatment may fail in some patients who may then respond to azathioprine and 6-MP. Aside from increased susceptibility to infections, the other side effects of azathioprine and 6-MP include hepatitis, pancreatitis and bone marrow depression.

Thiopurine methyltransferase (TPMT) levels are important in the safe use of azathioprine and 6-MP. Azathioprine is converted into 6-MP in the body and 6-MP then is partially inactivated by TPMT before elimination. The activity of TPMT enzyme is however genetically determined, and approximately 10% of individuals have reduced or absent TPMT activity. These patients thus accumulate 6-MP with consequent risk of bone marrow toxicity manifested by leucopenia and reduced resistance to life-threatening infections. TPMT levels must be checked in each patient prior to initiation of treatment with azathioprine or 6-MP. Patients with low or absent TPMT activity are best treated with other drugs or prescribed substantially low doses of 6-MP or azathioprine and checked frequently for marrow depression. Additionally allopurinol must not be prescribed together with azathioprine or 6-MP as it reduces TPMT activity.

The combination of azathioprine and 6-MP in low doses (as distinct from the larger doses used in preventing rejection of transplanted organs) is effective in patients with moderate to severe Crohn's disease and ulcerative colitis with documented efficacy in 70% of patients. Azathioprine and 6-MP are indicated:

- in severe Crohn's disease and ulcerative colitis not responding to corticosteroids
- in patients who develop adverse effects on corticosteroid therapy
- in patients who develop corticosteroid dependency
- to maintain remission after surgical excision for complicated disease as an alternative to 5-ASA
- in patients with failed healing of perianal fistulas after treatment with metronidazole.

Anti-TNF-α agents

Infliximab (Remicade) is a humanized mouse antibody that binds to TNF-α, one of the cytokines produced by immune cells during activation of the immune system. TNF-α then stimulates other immune cells to initiate the cytokine cascade involved in the inflammatory response. Crohn's disease is characterized by excess continued production of TNF-α. Hence, infliximab, which binds and blocks the action of TNF-α, abrogates the inflammation. Infliximab is used with success (65% response) in the treatment of moderate to severe Crohn's disease patients who respond inadequately to corticosteroids, azathioprine or 6-MP. Improvement is observed within days in some patients and within 2 weeks in the majority after commencement of the drug, which is administered by intravenous infusion. Infliximab is also effective in the healing of resistant perianal fistulas. Repeated infusions of infliximab every 8 weeks are used to maintain remission in many patients over a 1–2 year period. Response to infliximab after repeated infusions however declines and the drug may cease to be effective due to the development of antibodies to the infliximab.

Currently, infliximab is used for fast symptomatic control of active and severe disease when it is stopped and followed by maintenance treatment with azathioprine, 6-MP or 5-ASA compounds. There is some evidence that azathioprine or 6-MP may also help in preventing or delaying the development of antibodies against infliximab. Although generally well tolerated, adverse effects may occur during infliximab infusions. These include chest pain, shortness of breath, nausea and headaches. There is also an increased risk of infection including pneumonia and tuberculosis (TB). Hepatosplenic T-cell lymphoma has been reported in patients with Crohn's disease on azathioprine therapy alone or in combination with infliximab. Additionally, patients also may develop a delayed allergic reaction 7–10 days after receiving the infliximab with flu-like symptoms (fever, joint pain and swelling) and a worsening of Crohn's disease activity index.

The current recommended treatment with infliximab consists of a three-dose cycle at time zero, 2 weeks and 4 weeks to induce a remission with maintenance infusions every other month or 6-MP or 5-ASA compound. Infliximab is contraindicated in patients with existing infections, e.g. pneumonia, urinary tract infections, or abscess formation. In view of the reported cases of TB, it is now recommended that patients should be tuberculin tested before receiving infliximab.

Adalimumab (Humira) is another anti-TNF-α agent with similar action to infliximab but is administered subcutaneously. Several clinical trials have confirmed its efficacy in reducing signs and symptoms of Crohn's disease. It is also helpful in treating arthritis and arthralgias in patients with moderate to severe active Crohn's disease. In other respects, therapy with adalimumab provides equivalent results to infliximab in inducing and maintaining remission in patients suffering from Crohn's disease. Like infliximab, it is also effective in healing perianal fistulas. It is also effective in patients who do not respond or who are intolerant to infliximab. In adult patients adalimumab is administered subcutaneously every 2 weeks.

The most common side effects of adalimumab are skin reactions at sites of injection with swelling, itching and erythema. Other common side effects include upper respiratory infections, sinusitis and nausea. As with infliximab, adalimumab increases the risk of infection and the same precautions are needed including tuberculin testing prior to commencing treatment. It is also contraindicated in established infections. There have been reports of deteriorating cardiac function and heart failure in patients on infliximab or adalimumab.

Certolizumab pegol (Cimzia) is a pegylated humanized antibody fragment also directed against TNF-α with similar actions to infliximab and adalimumab. Although unlike these monoclonal antibodies certolizumab is an incomplete antibody and, therefore, does not induce *in vitro* complement activation. It is used in patients with moderate to severe Crohn's disease who do not respond adequately to standard medical therapy. Certolizumab is effective in inducing and maintaining clinical remissions for over 3.5 years. The standard dose of certolizumab is 400 mg subcutaneously at week 0, week 2 and week 4 to induce remission. Maintenance dosing is with 400 mg subcutaneously every 4 weeks.

Certolizumab is well tolerated. The most common side effects necessitating cessation of treatment include abdominal pain and diarrhoea. Intestinal obstruction has been reported but is rare. Hypersensitivity reactions have also been reported including angioedema, allergic dermatitis, shortness of breath, hot flushes, hypotension, malaise and syncope.

Anti-α-4 integrin

Natalizumab (Tysabri) is a humanized monoclonal antibody to alpha-4 integrin, also effective in patients with moderate to severe Crohn's disease unresponsive to aminosalicylates, antibiotics, corticosteroids, immunomodulators, or TNF-α inhibitors. It targets the cellular adhesion molecule α-4-integrin, which is expressed on leucocytes involved in the pathogenesis of Crohn's disease. The recommended dose is 300 mg by intravenous infusion over 1 hour every 4 weeks. It must not be used in conjunction with immunosuppressants or TNF-α inhibitors. The most common adverse effects are headache, fatigue, upper respiratory infections and nausea. The most serious adverse events are hypersensitivity, immunosuppression/infections and progressive multifocal leucoencephalopathy (PML), which is caused by reactivation of a latent human JV polyoma virus and is usually fatal.

Methotrexate

This acts as both an immunodepressant and anti-inflammatory agent and has been in established use for several years in treating patients with moderate to severe Crohn's disease who are either resistant or intolerant to azathioprine and 6-MP. It is administered either orally or by weekly subcutaneous or intramuscular injection. When used over long periods (years) it may cause hepatic cirrhosis. This risk is higher in patients who drink alcohol in excess and in severely obese patients. Other side effects of methotrexate include low leucopenia and pulmonary fibrosis. Because of potential damage to the foetus, methotrexate is contraindicated in pregnancy.

The Second European evidence-based consensus recommends management protocols which depend on disease severity (Box 29.1).

BOX 29.1 Second European consensus on evidence-based management of Crohn's disease

Mild to moderate active disease

- Oral mesalamine 3-4 g daily or sulfasalazine for ileocolonic or colonic disease as 3–6 g daily in divided doses
- Budesonide (9 mg/day) for disease confined to the ileum and/or right colon
- Proton pump inhibitors in patients with upper gastrointestinal Crohn's disease

Moderate to severe disease

- Prednisone 40–60 mg/day until resolution of symptoms
- Appropriate antibiotic therapy for infection or abscess
- Azathioprine and 6-MP for maintaining a steroid-induced remission
- Methotrexate 25 mg/week for steroid-dependent and steroid-refractory Crohn's disease
- Infliximab, adalimumab, and certolizumab pegol for moderate to severe disease in patients who do responded to steroid or other immunosuppressive drugs
- Natalizumab for patients with moderate to severe disease who do not obtain an adequate response or are unable to tolerate conventional Crohn disease therapy and anti-TNF antibody therapy

Perianal or fistulizing disease

- Surgical drainage of abscesses
- Treatment with antibiotics (metronidazole), immunosuppressives, or infliximab

Surgical treatment

Owing to the chronic relapsing nature of Crohn's disease surgical treatment can never be regarded as curative but it has an extremely important role to play; and all patients with Crohn's disease should have the benefit of a combined medical and surgical approach. Surgical intervention is indicated for the complications of the disease and for active disease which has not responded to medical therapy. However, there is considerable morbidity associated with the surgical treatment and about half of the deaths of patients with Crohn's disease are associated with operative intervention. It has to be appreciated, however, that patients coming to surgery for Crohn's disease often have serious complications of the disease.

Surgical treatment is indicated in Crohn's disease only for:

- excision of a diseased segment causing small bowel obstruction
- treatment of perforation
- drainage of intra-abdominal or perirectal abscesses
- treatment of severe perianal and anal fistulas unresponsive to medical treatment
- resection of symptomatic internal fistulas
- complications of medical management in particular in patients with major side effects of steroid therapy
- growth retardation in children
- development of obstructive uropathy
- massive haemorrhage (rare).

The resection must aim to conserve as much as is possible of the small bowel and for this reason must be strictly limited to the diseased segment(s). In many patients, however, Crohn's disease eventually recurs, affecting previously healthy bowel. As recurrence is common after resection (50% within 4 years of surgery), treatment with drugs such as mesalamine, Pentasa or 6-MP is recommended to reduce the risk of relapse. Infliximab has also been shown to be effective in reducing postoperative recurrence after ileocaecal resection, although relapse may occur when this treatment is stopped.

Other factors which influence the decision to operate and the type of operative procedure include the anatomical site of the disease, the surgeon's preference and expertise and whether the operation is being performed as an emergency or as an elective procedure. The extra gastrointestinal manifestations of Crohn's disease are rarely, if ever, an indication for operative intervention, and certainly in small bowel disease resection does not appear to have a major effect in this respect.

When making an abdominal incision it should be planned with attention paid to previous incisions and the siting of a stoma. Laparoscopic assisted surgery is appropriate for patients with isolated terminal ileal or ileocolic disease where a limited right hemicolectomy is required and several such reported series have documented the benefit of the laparoscopic approach in these patients. Many patients, however, still require open surgery. At laparotomy, an initial exploration is carried out to determine the extent and severity of the disease and to establish the state of the liver and the presence and otherwise of gall stones. Adhesions which are often present are taken down and the anatomy and the length of the small intestine ascertained.

The emphasis of modern surgical treatment for small bowel disease is on the maximal preservation of intestine with resections being strictly limited to the diseased segment causing the symptom or complication. Bypass should only be carried out when there is no other alternative, as blind loops encourage bacterial overgrowth, increased incidence of abscess and fistula formation and risk of developing carcinoma. In deciding the extent of small bowel resection a balance must be struck between removing all grossly diseased bowel and retaining sufficient for adequate absorption and nutritional support. Macroscopic assessment of the disease is used to determine the extent of the resection. The important macroscopic features in this respect include thickening of the bowel wall and mesentery, hyperaemia and oedema, fat encroachment of the bowel and the presence of ulcers at the resection line. It is now established that areas of minimal disease may be safely left as there is no evidence that microscopic disease at the resection margin reduces the recurrence rate. Skip lesions in the vicinity of the main disease process may be safely excised *en bloc* but those at a distance should be resected separately if considered significant.

Conservative (non-resectional) surgery is particularly indicated in patients with multiple previous resections and in those with multiple strictures. In order to determine whether or not a small area of disease is causing a significant stricture, it is useful to manipulate the balloon of a Fogarty catheter along the lumen of the bowel introduced via an enterotomy. This will give some indication as to the size of the lumen at the strictured site. If the stricture admits the passage of an 18–20 mm inflated balloon, intervention is probably not necessary.

Stricturoplasty is now widely used for the treatment of strictures. The type of procedure used depends on the length of the stricture, a Heineke–Mikulicz stricturoplasty being ideal for short stenotic areas (Figure 29.27) and the Finney equivalent for long strictures (Figure 29.28). The decision between a stricturoplasty and a resection for a long stenotic segment is dependent on surgical experience and expertise and it has to be remembered that stricturoplasty is associated with postoperative morbidity including a leak rate of around about 10%. Although there are no prospective randomized trials the immediate outcome of stricturoplasty appears to be similar to that encountered after resection for small bowel Crohn's disease and there is no evidence that there is a significant increased risk of recurrence after stricturoplasty.

Compared to resection, stricturoplasty is not appropriate in the presence of sepsis (abscess) and fistula, both of which are best treated by resection. Adhesions and fistulas between diseased segments of bowel and adjacent hollow viscus are dealt with by careful separation of the diseased segment from the adjacent structure with debridement and primary closure of the latter. External fistulas are managed by similar excision with debridement and closure of the abdominal wall but not of the skin and the subcutaneous tissues. Secondary wound closure is performed several days later. If there is a mass involving several loops of bowel the surgeon must carefully identify and dissect healthy from diseased bowel. Provided that the length involved is not great, a resection en mass may be performed. After the small bowel resections have been completed the length of the residual small intestine must be measured along the anti-mesenteric border and recorded.

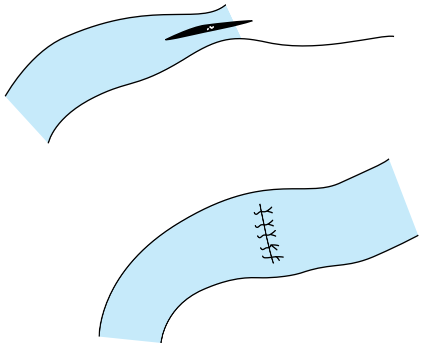


Figure 29.27 Diagrammatic representation of the technique of Heineke-Mikulicz stricturoplasty. This operation is suitable for short strictures in the absence of active inflammation. An operative biopsy with frozen section histology is advisable to exclude carcinoma.

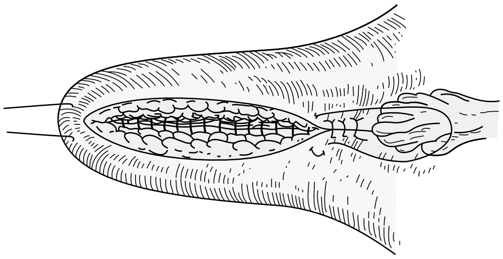


Figure 29.28 Diagrammatic representation of Finney-type stricturoplasty for long stenosed segments of Crohn's disease.

Free perforation into the peritoneal cavity is rare in Crohn's disease and the commonest site is the ileum. Resection of the diseased segment with peritoneal lavage and debridement is the treatment of choice. Primary anastomosis or exteriorization is performed depending on the duration and the severity of the peritoneal inflammation. Simple drainage of an abscess invariably results in a fistula. As described above, some 30% of enterocutaneous fistulas may heal with total parenteral nutrition but persistence of the fistula beyond 6 weeks is an indication for surgical intervention with excision of the diseased segment of bowel. Enterocolic and enterovesical fistulas never heal by conservative management and always require surgical intervention. Perforation of the diseased bowel into the retroperitoneal tissues causes intense fibrosis and retroperitoneal inflammation resulting in an obstructive neuropathy. Although the obstructed ureter may be asymptomatic the resulting renal impairment may be progressive and permanent. Hence, obstructive uropathy is an absolute indication for surgery and is best relieved by appropriate bowel resection and ureterolysis.

Operations for Crohn's disease are followed by a high incidence (10–15%) of postoperative complications such as anastomotic leakage, fistula formation, intra-abdominal abscess and haemorrhage. The overall reported mortality rate ranges from 2% to 8%. A high mortality is associated with emergency operations (up to 30%). Late postoperative complications include short bowel syndrome, urinary lithiasis, cholelithiasis, gastric hypersecretion and peptic ulcer disease.

One of the unresolved problems is the high incidence of recurrence following resection. Accumulative recurrence after resection of small intestinal Crohn's disease is 30% at 5 years, 50% at 10 years, and 60% at 15 years. Young patients and those with ileocolic disease have the highest and earliest recurrence and the worst ultimate prognosis. The prognosis of patients with Crohn's colitis undergoing proctocolectomy is better in this respect as the development of ileal disease occurs in 10% and 20% of these patients at 5 and 10 years respectively. After abdominal colectomy and ileorectal anastomosis the reported recurrence rate averages 50% at 5 years. The majority of recurrences occur at or near the anastomosis and in patients with ileorectal anastomosis. The recurrent disease often involves the ileum and further resection does not necessarily require proctectomy and ileostomy.

In assessing the widely divergent reported figures for recurrence rates after surgical intervention for Crohn's disease two points must be stressed: (1) the recurrence rate rises gradually with the duration of follow-up and (2) it is essential to differentiate patients with true recurrent Crohn's disease from those requiring reoperations for other complications including ileostomy revisions.

Many patients with Crohn's disease undergo multiple resections and a significant percentage of these develop the short gut syndrome. At present the majority of these patients are managed by home parenteral nutrition which may be complicated by catheter sepsis, venous thrombosis and parenchymal liver disease. Recently the development of more effective immunosuppressive drugs (cyclosporin especially) and the discovery of the protective effect of a concurrent liver transplant (especially on graft-vs-host disease) has led to the successful use of combined hepatic and small bowel transplantation. The early results have been encouraging but it has yet to become a routine procedure. One of the major concerns is the possibility of recurrence of Crohn's disease in the transplanted bowel but there are very little data to support this at present.

Polyarteritis

Pathology

This condition is characterized by a systemic necrotizing vasculitis with fibrinoid necrosis and an inflammatory cell infiltrate affecting the blood vessels (arteries and veins) of several organs including the small intestine. The weakened vessels lead to aneurysm formation, which may rupture and bleed or thrombose causing multiple organ infarcts. There are several categories of polyarteritis but the two which affect the gut most frequently are polyarteritis nodosa (PAN) and the *Churg–Strauss syndrome* (CSS) also known as allergic granulomatous angiitis because of the eosinophilic infiltration and granuloma formation in the connective tissue in addition to the vasculitis.

Aetiology

The pathogenesis of PAN is unknown but, in some cases, PAN is associated with viral infections, especially hepatitis B virus (HBV) (10–15%). Evidence for immune complex-induced disease is thus confined to HBV-related PAN and the role of immune complexes in non-HBV-related PAN remains unclear. Impaired function of endothelial cells may be part of idiopathic PAN or a consequence of it.

CSS is a granulomatous small-vessel vasculitis of unknown cause. There is no firm evidence for the role of immune complexes or cell-mediated mechanisms in CSS, although autoimmunity is involved as indicated with the presence of hypergammaglobulinaemia, increased levels of IgE and rheumatoid factor. A CSS-like syndrome may develop in patients with asthma who are steroid-dependent on treatment with leukotriene receptor antagonists on reduction of the oral steroid dose. The steroid withdrawal is thought to unmask the syndrome although patients have been documented who developed vasculitis when a leukotriene receptor antagonist was substituted for inhaled steroids in the absence of oral steroid withdrawal.

Clinical features

The American College of Rheumatology (ACR) recommends the presence of at least three out of the following criteria for establishing a clinical diagnosis of PAN. Additionally the diagnosis should be confirmed by biopsy:

- livido reticularis (mottled purplish discoloration of the skin of the torso and extremities)
- testicular pain
- muscle pain, weakness and tenderness
- neuropathy (single or multiple)
- hypertension (systolic >90 mmHg)
- impaired renal function (blood urea nitrogen >40 mg/dL or creatinine >1.5 mg/dL)
- positive HBV test for surface antigen or antibody
- angiograpy showing vascular damage (aneurysms or stenosis)
- biopsy showing arteritis (necessary for confirmation).

The ACR also recommends the following criteria for the diagnosis of CSS. The presence of four or more criteria yields a sensitivity of 85% and a specificity of 99.7%:

- asthma (wheezing, expiratory rhonchi)
- eosinophilia of more than 10% in peripheral blood
- paranasal sinusitis
- pulmonary infiltrates (may be transient)
- histological proof of vasculitis with extravascular eosinophils
- mononeuritis multiplex or polyneuropathy.

In addition to systemic manifestations (fever, weight loss, malaise and hypertension) the symptoms emanating from involvement of the small intestine include nausea, vomiting, diarrhoea and steatorrhoea.

Complications

PAN and CSS can lead to life-threatening complications such as intestinal infarction, perforation and massive gastrointestinal haemorrhage.

Treatment

Medical treatment with corticosteroids and either cyclophosphamide or azathioprine is the mainstay of therapy. Surgery is only undertaken for the major complications which entails resection of the affected segment of small bowel. Particularly in cases of infarction the ends should be exteriorized rather than attempting a primary anastomosis with subsequent restoration of continuity once bowel viability is assured.

Plasma exchange is indicated in HBV-related PAN, where it is the most effective treatment after an initial short-tem course of prednisone and is administered in conjunction with antiviral medications, e.g. vidarabine and interferon-α. Plasma exchange is not effective and is therefore not indicated in PAN not associated with HBV infection and in CSS. Prednisone in combination with cyclophosphamide is the current recommended first-line treatment in these cases despite infectious side effects, which can be reduced by dose adjustment to the neutrophil and lymphocyte counts.

Eosinophilic gastroenteritis

Eosinophilic gastroenteritis (EGE) is an uncommon gastrointestinal disease affecting both children and adults which is characterized by (1) the presence of gastrointestinal symptoms which may be chronic or acute, (2) eosinophilic infiltration in one or more areas of the gastrointestinal tract, defined as 20 or more eosinophils per high-power field in the absence of an identified cause for the eosinophilia and without involvement of extra gastrointestinal tissues.

Aetiology and pathology

The aetiology and underlying molecular mechanisms responsible for EGE are not known although recent studies implicate the role of esosinophils, T helper 2 and chemokines (eotaxin) in the pathogenesis. Eosinophils function as antigen-presenting cells and express major histocompatibility complex class II molecules. They are known to mediate proinflammatory effects, including the upregulation of adhesion molecules, modulating cell trafficking and inducing cell activation by releasing various cytokines and transforming growth factor-α/β, chemokines (RANTES and eotaxin) and lipid mediators (platelet-activating factor and leukotriene C4). Additionally, eosinophils can induce tissue damage by releasing toxic granule proteins, e.g. eosinophilic cationic protein, eosinophil peroxidase, and eosinophil-derived neurotoxin.

The disease may manifest as eosinophilic gastritis or enteritis. Atopy is present in a subset of patients, as demonstrated by raised total IgE on food-specific IgE radioallergosorbent assay test (RAST) or skin tests. Macroscopically the wall of the affected segment is thickened and there are small mucosal ulcers.

Clinical features

The clinical manifestations vary with the site of involvement in the gastrointestinal tract and the extent of histological involvement of the gut wall: mucosal, muscularis or serosal disease. Patients present usually in the third to fifth decades of life, but the disease can affect any age group, including infants. The disease most often involves the stomach and the small bowel. The mucosal form of EGE is characterized by vomiting, dyspepsia, abdominal pain, diarrhoea, faecal blood loss, iron-deficiency anaemia, malabsorption, protein-losing enteropathy and failure to thrive in children. The muscularis form may present with gastrointestinal obstructive symptoms mimicking pyloric stenosis or intestinal obstruction. The less common serosal form presents with significant bloating, ascites and very high peripheral eosinophilia.

Diagnosis

Diagnosis is based on history, clinical picture, elevated IgE, endoscopy and biopsy and elevated peripheral eosinophil counts. Some cases presenting with pyloric or intestinal obstruction are first diagnosed at laparotomy. The raised blood eosinophil count should alert the clinician to the possibility of EGE.

Treatment

The treatment of EGE is unsatisfactory. When a short segment of diseased small bowel is encountered at laparotomy for intestinal obstruction, the segment is usually resected and the diagnosis established by the subsequent histology. However, surgery is not advisable if the diagnosis is made preoperatively; treatment is medical with corticosteroid therapy and surgery is reserved for patients in whom the obstructive symptoms persist.

In addition to oral steroids or cromolyn, patients with mucosal disease may also benefit from diet elimination therapy, particularly those with a history of food intolerance or allergy.

Radiation enteropathy

Pathology

The immediate effect of radiation on the gastrointestinal tract is arrest of cell division in the intestinal crypts. This effect is largely restricted to cells in the G1 phase. With greater radiation exposure, oedema and ulceration of the mucosa ensue. As a result of the diminished cell turnover the mucosa becomes thinner with stunted villi. If there is extensive small intestinal involvement, malabsorption of varying degrees will occur.

The long-term effects are due to transmural fibrosis following the appearance of atypical fibroblasts in the submucosa and ischaemia from a proliferative endarteritis and vasculitis.

Obliteration of the intestinal lymphatics with lymphatic ectasia complicates the pathological picture. The oedema is most marked in the submucosal layer. The fibrosis which affects all the coats is progressive and accompanied by hyalinization. The muscle layers exhibit areas of myofibrillar degeneration with atrophy of the muscle fibres, patchy hyalinization and disturbed motility. The serosa becomes thickened, opaque and greyish white and dense adhesions develop between adjacent intestinal loops.

Incidence

Although the figure of 5% is often quoted for the incidence of intestinal radiation-induced bowel disease in patients who receive radiotherapy to the abdominal and pelvic regions, the individual estimates vary from 3% to 25%. There is however general agreement that the prevalence is probably underestimated due to lack of clinical recognition. The cumulative 10 year morbidity after abdominopelvic radiotherapy is estimated at 8%, with severe enteropathy accompanied by major complications (bleeding and obstruction, stenosis and fistula formation, malabsorption and perforation with peritonitis) in 3%. The most common situation is where the pelvis is irradiated, usually for rectal or gynaecological cancer. In the normal situation and particularly after rectal excision, loops of small bowel lie in the pelvis and are at great risk of irradiation.

Clinical features

Symptoms are encountered in the majority of patients during the first few weeks of radiotherapy. Anorexia, nausea and vomiting are CNS effects as these are often encountered in patients receiving radiotherapy in extra-abdominal regions. These early symptoms usually subside rapidly and do not necessarily indicate the development of late sequelae that characterize radiation-induced bowel disease.

The commonest symptoms referable to chronic bowel damage are vague abdominal discomfort, diarrhoea, mild rectal bleeding and the passage of mucus. The interval between the time of radiation and onset of symptoms varies considerably from 2 months to 2 years. Intestinal obstruction may be acute or subacute and recurrent. Occasionally, acute presentation with infarction may occur and this carries a high risk of perforation and mortality. Most of the serious late complications tend to occur within 2 years of the initial treatment but may become progressively worse with time.

Diagnosis

The investigative procedures used in the assessment of patients include contrast radiology of the small and large intestine and malabsorption studies. Sigmoidoscopy and colonoscopy may be appropriate in patients with large bowel disease.

Complications

As indicated under Clinical features, the main complications are intestinal obstruction and perforation secondary to infarction (Figure 29.29). Fistula formation can also occur.

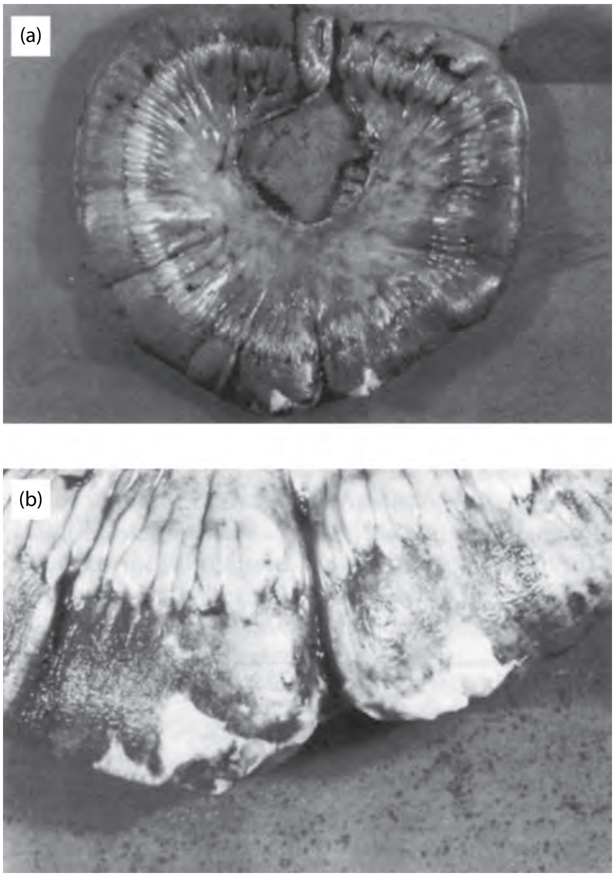


Figure 29.29 (a) Emergency resection of infarcted radiation enteropathy of the small bowel; (b) close-up of the perforation.

Treatment

Prophylaxis

This is of great practical importance and includes medical, radiotherapeutic and surgical measures.

The preventive medical measures designed to decrease the effects of radiation on the gastrointestinal tract include:

- maintenance of adequate hydration
- sulfasalazine, 500 mg orally twice daily, reduces the incidence and severity of radiation-induced enteropathy in patients receiving external beam radiotherapy
- amifostine (a prodrug which is converted to its intracellular metabolite, WR-1065), which acts as a radioprotectant. Oxidation of WR-1065 produces metabolites which act as scavengers of free radicals, thus minimizing DNA damage. Amifostine administered intravenously (340 mg/m^2) has been shown to decrease the incidence of radiation proctitis in patients receiving radiotherapy for rectal cancer
- sucralfate, orally and as enemas, is used as a mucosal protective agent in the prophylaxis against radiation enteropathy, with variable results
- glutamine, which is the preferred metabolic fuel for enterocytes, also has antioxidant function and is used both in the prophylaxis and in the treatment of established disease
- hyperbaric oxygen therapy is also used in the prophylaxis of radiation proctitis.

The various techniques used to minimize radiation-induced damage to the gut in patients requiring abdominopelvic external radiotherapy are:

- multidirectional, sharply collimated beams
- multiple-field, conformal therapy with prior three-dimensional planning
- computer-assisted dosimetry
- more stable intracavitary applicators
- extended intervals between fractionated doses and a lower dose per fraction.

Surgical measures are used in patients undergoing surgery who will require postoperative radiotherapy. They are based on the principle of reperitonealization and abdominopelvic partitioning and are aimed at excluding the mobile small bowel from the pelvic cavity where the small bowel loops become adherent to each other and to the pelvic wall. Both native tissue and prosthetic material are used to create a partition between the abdominal and pelvic cavities or to retain the small bowel within the abdominal cavity after surgery. In reperitonealization procedures, native tissue (peritoneum, omentum, urinary bladder, uterine broad ligaments) are used for the partitioning of the two pelvic cavities. In omental-based procedures, an omental flap based on the left gastroepiploic pedicle is used to retain the small bowel within the abdominal cavity. Prosthetic procedures include the use of absorbable mesh slings and less commonly temporary pelvic space-occupying devices such as inflatable balloons or silicone implants which have to be removed after the completion of radiotherapy.

Medical therapy of radiation enteropathy

A low-fat, low-residue, lactose-free diet (many patients become intolerant of lactose) is recommended, with octreotide being reserved for patients who do not obtain symptomatic relief. Nausea is treated with antiemetics. Other agents of marginal value include 5-ASA and sucralfate as a mucosal protective agent. Iron supplementation is administered in patients with low-grade bleeding leading to iron-deficiency anaemia. Topical steroids and sucralfate enemas are indicated for chronic radiation proctitis. The presence of significant malabsorption necessitates further measures, such as antibiotics for bacterial overgrowth and bile salt binding agents (cholestyramine) for ileal disease together with careful dietary management, the use of elemental diets and oral supplements. Glutamine-supplemented elemental diets are needed in patients with severe malabsorption. Some patients with extensive small bowel disease and severe malabsorption require parenteral nutrition intermittently or indefinitely.

Surgical treatment of radiation enteropathy

Whenever possible, management should be conservative as surgical treatment has a high morbidity and an average reported mortality of 25%. Emergency surgery is required for infarction with perforation or acute intestinal obstruction which does not resolve on conservative treatment. The radionecrotic bowel is ideally excised with primary anastomosis or exteriorization of the bowel ends in the presence of ischaemia and sepsis. Occasionally when the small bowel is firmly adherent within the pelvis, dissection of the affected segments may be hazardous and ileoileal or ileocolonic bypass may be more appropriate. Elective surgery may be undertaken for chronic severe symptoms due to stricture or internal fistula formation and the same considerations apply. In the performance of these operations only these adhesions which are in the operative field should be divided. Extensive adhesiolysis should be avoided as this is hazardous and can be complicated by perforation.

Infective conditions of the small bowel

Campylobacter gastroenteritis

Campylobacter species are among the most common pathogens in humans and are commensal in many animals (chickens, other birds, swine and cattle). The organisms constitute the most common cause of culture-proven bacterial gastroenteritis/colitis in both developed and developing countries and the incidence in the number of cases often exceeds those of salmonellosis and shigellosis. Thus these infections have become a major public health problem and, according to the WHO, *Campylobacter* genus is one of the most frequently isolated bacteria from stools of infants with diarrhoea in developing countries.

Clinical features

Whereas in developing countries *Campylobacter* gastroenteritis has no seasonal incidence, epidemics have been well documented during summer and autumn in developed countries. Although *Campylobacter* gastroenteritis/colitis is usually a mild self-limiting disorder with diarrhoea and colicky abdominal pain, it may rarely cause a severe infection associated with bacteraemia, abscess formation and other internal abdominal infections. When severe, the disease may also be associated with the development of the Guillain–Barré syndrome (acute inflammatory demyelinating polyneuropathy).

The diagnosis is made on stool culture but occasionally because of undiagnosed severe abdominal pain and tenderness a laparotomy may be performed usually with negative results. However, there are rare instances of *Campylobacter* appendicitis, cholecystitis and pancreatitis.

Treatment

Rehydration and correction of electrolyte abnormalities constitute the treatment. Debate exists on the need for antibiotic therapy. This is certainly not indicated in patients with mild disease and should only be considered for patients with severe symptoms, fever and bloody diarrhoea, and in immunosuppressed patients. *Campylobacter* species are often resistant to several antibiotics including penicillin, ampicillin and cephalosporins. Although most strains are susceptible to erythromycin, azithromycin, gentamicin and tetracycline, resistance may develop. When antibiotic therapy is indicated, erythromycin for 5–7 days is considered the appropriate treatment. Except in the patients with severe disease, there is considerable debate on the benefit of antibiotic therapy especially as a significant therapeutic effect is only observed when treatment is commenced early during the course of illness. There are very few randomized studies that compare antibiotics with placebo for the treatment of *Campylobacter* species-related diarrhoea. Some of these studies have concluded that antibiotics have no effect; but a meta-analysis showed a significant decrease in symptoms and bacterial shedding time together with a trend towards a beneficial effect when early antibiotic treatment was started.

Yersinia infections

These Gram-negative coccobacillary rods belong to the family Enterobacteriaceae. The organisms belong to the genus *Yersinia,* which includes 11 species, only three of which are pathogenic to humans: *Y. pestis* (bubonic plague), *Y. pseudotuberculosis* (mesenteric adenitis), and *Y. enterocolitica*. *Yersinia* species are Gram-negative, oxidase-negative, and non-lactose-fermenting bacilli.

The two species which cause gastrointestinal infections particularly of the terminal ileum, appendix, ascending colon and mesenteric lymph nodes are *Y. enterocolitica* and *Y. pseudotuberculosis*. The infection induces a granulomatous inflammatory picture with microscopic abscess formation which may simulate chronic inflammatory bowel disease. The ileum and ascending colon may become swollen, ulcerated and exhibit macroscopic appearances similar to Crohn's disease.

Pathology

Transmission is primarily via ingestion of contaminated foods, including pasteurized milk and dairy products as *Yersinia* can proliferate at refrigerated temperatures. There are 34 serotypes of O antigen of *Y. enterocolitica*, types 0:3 (sporadic cases); 0:5,27; 0:8 (foodborne outbreak); and 0:9 being the most common and virulent types. *Y. enterocolitica* grows in environments with a low pH 5–9, hence the increased incidence of the disease in patients who take antacids and acid suppression drugs. Also, the organisms require iron to multiply, and this accounts for the reported cases in patients with iron overload (haemochromatosis).

After ingestion the bacteria reach the terminal ileum which is the site of mucosal adherence and penetration. This is followed by bacterial proliferation in the Peyer's patches followed by the development of non-specific ileocolitis with an inflammatory infiltrate in the lamina propria (usually limited to the right colon). The bacteria may then spread to the mesenteric lymph nodes, which may then cause bacteraemia with or without the development of intra-abdominal abscesses and pain in the right lower quadrant that mimics acute appendicitis. Although *Y. enterocolitica* produces enterotoxin, this does not contribute to the pathogenicity which is due to plasmid-encoded proteins of the outer membrane of the organism, responsible for the mucosal adherence and invasion. Antibodies directed against these proteins are present in patients convalescing from disease.

Diagnosis

The diagnosis is established by recovery of the organism from the stool. However, in patients with chronic disease, radiological investigation may show nodular filling defects in the terminal ileum which may simulate Crohn's disease. When operation is carried out the terminal ileum and mesenteric lymph nodes are found to be inflamed and swollen. Biopsy of the lymph nodes for *Yersinia* may be performed with safety in these patients, but an appendicectomy is ill advised.

Clinical features

Y. enterocolitica infections are more common in cold climates and hence prevalent in northern Europe. The incubation period is 1–14 days but the duration of stool excretion of the organisms is much longer (up to 3 months). The majority (75%) of patients are children aged 5–15 years. The spectrum of disease ranges from asymptomatic to an acute self-limiting gastroenteritis to life-threatening sepsis especially in infants.

The most usual clinical syndrome is an acute febrile gastroenteritis which is self-limiting lasting 5–14 days. However, chronic symptoms including diarrhoea and rectal bleeding may occur that are associated with persistent inflammation, particularly in children. Several syndromes are associated with *Yersinia* infection in children. These include enterocolitis, the pseudoappendicitis syndrome, bacteraemia and postinfectious sequelae.

Enterocolitis

This is the most common presentation and occurs primarily in young children (mean age of 24 months). The incubation period is 4–6 days, with a range of 1–14 days. The prodromal symptoms include listlessness, anorexia and headache which are followed by watery, mucoid diarrhoea, fever, colicky abdominal pain and bloody stools. The diarrhoea may be short lived but may also last a few weeks. Concomitant bacteraemia occurs in 20–30% of infants younger than 3 months. Most cases are self-limited. Complications include diffuse ulceration and inflammation of the small intestine and colon, peritonitis, meningitis, intussusception and cholangitis.

Pseudoappendicitis syndrome

This is characterized by fever, abdominal pain, tenderness in the right lower quadrant and leucocytosis and is usually caused by *Y. pseudotuberculosis*. The infection causes mesenteric lymphadenitis with terminal ileitis. The pseudoappendicitis syndrome is more common in older children and young adults.

Bacteraemia

Bacteraemia is seen most frequently in very young infants and patients with iron-overload syndromes receiving frequent transfusion (sickle cell anaemia, thalassaemia) and in patients on oral iron supplements.

Postinfectious, non-suppurative sequelae

Although uncommon, these may cause considerable morbidity. They include reactive polyarthritis, erythema nodosum and proliferative glomerulonephritis. They are usually encountered in individuals with HLA-B27. Chronic inflammatory bowel disease has also been associated with *Y. enterocolitica* infections.

Treatment

In mild disease, treatment is primarily supportive with fluid and electrolyte therapy. Antibiotics do not influence the course of uncomplicated enteritis. However infants younger than 3 months and immunocompromised children require aggressive in-hospital treatment with intravenous antibiotics. Tetracycline has been traditionally the drug of choice. However, with the increasing resistance to this antibiotic and its inadvisability in children younger than 8 years, alternative first-line drugs which include aminoglycosides and trimethoprim-sulfamethoxazole are used nowadays. Other effective drugs include third-generation cephalosporins and chloramphenicol. Antimotility agents are contraindicated in the treatment of *Y. enterocolitica* infection because of the increased risk of invasion.

Typhoid fever (enteric fever)

Typhoid fever (enteric fever) is a serious potentially life-threatening multisystemic infection caused primarily by *Salmonella typhi*. The disease is prevalent in conditions of poor sanitation and overcrowding and is endemic in developing countries. Typhoid fever is characterized by an early septicaemic phase with colonization of several organs, i.e. liver, spleen, bones and small intestine. The classic presentation is with fever, malaise, diffuse abdominal pain and constipation. However, the disease is known for its protean manifestations, such that it often poses a diagnostic challenge. Survivors may be left with long-term or permanent neuropsychiatric complications or a chronic carrier state. The terminal ileum in the region of the Peyer's patches is the commonest site for intestinal infection with the formation of longitudinal ulcer(s) on the antimesenteric border situated within 45 cm of the ileocaecal valve in the majority of patients.

Pathophysiology

Typhoid fever is transmitted by the oral route (contaminated food including chicken and shell fish and drinks) or because of poor personal hygiene resulting in hand-to-mouth transmission following contamination from toilet seats. Whatever the route, only a small inoculum (100 000 organisms) is required for contracting the disease by this highly pathogenic organism.

Following ingestion non-typhoid *Salmonella* organisms are promptly phagocytosed by luminal macrophages, which then pass them through the intestinal mucosa. The macrophages recognize the pathogen-associated molecular patterns (PAMPs) and lipopolysaccharides by their special receptors and are thus able to recruit (by the release of IL-8) T-cells and neutrophils to mount an inflammatory response designed to protect the host. The situation is different for *S. typhi*, which enters the host primarily through the distal ileum using specialized fimbriae that adhere to the mucosal epithelium covering the Peyer's patches which serve as the main staging point for macrophages carrying *S. typhi* entering from the gut lumen. However, *S. typhi* has a Vi capsular antigen that masks PAMPs, avoiding the host defensive neutrophil-based inflammatory response. The organism is thus able to replicate within the macrophages as these reach the mesenteric lymph nodes and then through the thoracic duct and other lymph channels to the reticuloendothelial tissues of the liver, spleen, bone marrow and lymph nodes. Once there, the *S. typhi* bacteria continue to multiply to reach a critical density when they induce macrophage apoptosis, and are thus released and able to invade the bloodstream (bacteraemia) and establish the full-blown disease. The gallbladder is infected either by the bacteraemia or direct by extension from infected bile. *S. typhi* is thus able to re-enter the gut to reinfect the Peyer's patches. Some bacteria are shed in the stool and become a source for infection of other hosts.

Clinical features

The disease occurs sporadically in Western countries but is endemic in Asia, Africa, Latin America, the Caribbean, Oceania, Bangladesh, China, India, Indonesia, Laos, Nepal, Pakistan, and Vietnam. Within these countries, typhoid fever is prevalent in poor underdeveloped regions. Worldwide, the majority of typhoid fever cases involve school-aged children and young adults, although no age group is exempt. In children, the clinical picture may be atypical, ranging from a mild febrile illness to severe convulsions.

The classical presentation of typhoid fever after an incubation period of 7–14 days begins with a rising 'step-wise' temperature during the day that remits by the subsequent morning, although in recent years this remittent temperature is encountered less often. The patients then develop diffuse abdominal pain and tenderness during the first week of the illness. In some patients, the pain is diffuse, severe and colicky; in others it is constant and located in the right upper quadrant. The swelling of the Peyer's patches due to the inflammation narrows the bowel lumen, and is thought to be the cause of the constipation that characterizes the classical presentation of typhoid fever. As the disease progresses other symptoms develop, including dry cough, malaise, frontal headache, diminished consciousness and even delirium. The fever reaches a plateau of 39–40°C at the end of the first week, when typical pink 'rose spots' develop over the trunk but are usually few in number. They are caused by dermal bacterial emboli and resolve within a few days.

Abdominal distension and splenomegaly develop in the second week. At this stage, some patients may develop a bradycardia and a dicrotic pulse; they remain febrile and become more toxic and anorexic during the third week. Respiratory signs include tachypnoea and crepitations over the lung bases. Abdominal distension increases and may be accompanied in some patients by foul smelling, green liquid diarrhoea (pea soup diarrhoea) as the patient deteriorates to the 'typhoid state', characterized by apathy, confusion, and delirium. Rectal bleeding may also occur. Death is caused by overwhelming toxaemia and myocarditis with sudden death. A percentage of patients who survive the disease become *S. typhi* carriers forming a potential source of transmission of the disease to others, although asymptomatic carriers of *S. typhi* who have not had the disease are well documented.

The above classical picture may be absent. In the first instance, the 'stepladder' fever pattern previously considered the hallmark of typhoid fever is rarely encountered nowadays. Instead, the fever has an insidious onset and is steady. Also children and immunosuppressed individuals develop diarrhoea rather than constipation. Other atypical manifestations include severe headaches mimicking meningitis, acute lobar pneumonia, isolated infections, severe jaundice or pyrexia of unknown origin. In some countries, such as India and Africa, patients often present with neurological manifestations, e.g. delirium, parkinsonian symptoms or Guillain–Barré syndrome. When perforation of the terminal ileum occurs, the patient will develop severe lower abdominal pain and will have obvious signs of peritonitis on examination.

Diagnosis

Diagnosis is based on identifying the organism which can usually be found in faeces and urine in the second week. During the first week, blood culture and bone marrow culture may be positive. Liver function tests are often abnormal and, if perforation has occurred, a chest radiograph may show gas under the diaphragm. Serology (Widal's reaction) is of little value in the tropics as antibodies will be present from previous clinical or subclinical infection.

Complications

The main complication is perforation of the terminal ileum. This is solitary in 85% of cases and the incidence of perforation varies considerably from one endemic area to another. It is high in West Africa (15–33%) and low in Egypt and Iran (1–3%). The high incidence of perforation in West Africa has been attributed to late diagnosis and a particularly virulent strain of organism. The perforation is caused by necrosis of the Peyer's patches. Perforation is often unheralded and may be masked by corticosteroids (see Treatment). The other complications include intestinal haemorrhage and, rarely, pancreatitis, meningitis, orchitis, osteomyelitis and abscess formation anywhere on the body.

Treatment

The modern treatment of typhoid fever is governed by the following important considerations: (1) early diagnosis and antibiotic therapy curtails the severity of the disease and its course, (2) multidrug resistance to the antibiotics is an increasing problem and (3) the pattern of antibiotic resistance varies with the geographical area in which the disease is contracted. With prompt and appropriate antibiotic therapy (depending on best guess started before culture and sensitivity results become available), typhoid fever should be curtailed to a relatively short-term febrile illness requiring a median of 6 days of hospitalization, with few long-term sequelae and a low (0.2%) mortality.

Patients in Western countries with suggestive symptoms developing within 60 days of a visit to an overseas endemic area or those following consumption of contaminated food/drink should be started on the best-guess antibiotic therapy. The antibiotic regimen may need to be changed after the results of culture and sensitivity become available.

Previously healthy patients who remain in good condition with uncomplicated disease may be treated as outpatients but advised to use strict hand washing techniques and to avoid handling/preparing food for others. All other patients including children should be hospitalized and nursed in isolation during the acute phase of the infection with safe disposal of excreta (faeces and urine). Ill patients are best given intravenous fluids and electrolytes until the condition improves. Oral fluids and a soft low-residue digestible diet are commenced in the absence of ileus and distension.

Antibiotic therapy

Although the WHO guidelines recommend fluoroquinolone treatment for both complicated and uncomplicated cases of typhoid fever, most experts now consider these guidelines to be no longer applicable in view of the increasing resistance of *S. typhi* (up to 38%) in some countries to fluoroquinolones. The current recommendation is that the sensitivity pattern of *S. typhi* in the area of acquisition is the only valid criterion for choice of the antibiotic from the following, depending on the reported resistance in the area of acquisition of the infection:

- chloramphenicol (Chloromycetin) orally or intravenously
- amoxicillin administered orally
- trimethoprim and sulfamethoxazole administered orally
- ciprofloxacin and other quinolones (ofloxacin, norfloxacin, pefloxacin) [not currently recommended in children and pregnant women (induces cartilage damage)]
- gatifloxacin (third-generation fluoroquinolone)
- azithromycin (Zithromax) often used in children
- cephixime
- ceftriaxone (third-generation cephalosporin).

Chloramphenicol was used to treat typhoid until the 1970s, when widespread resistance developed. Ampicillin and trimethoprim-sulfamethoxazole were then used instead until the late 1980s, when some *S. typhi* and *S. paratyphi* strains developed plasmid-mediated resistance to all three antibiotics agents (multidrug resistant).

The third-generation cephalosporins have been used in regions with high fluoroquinolone resistance rates (Asia and Vietnam). Even so, in these countries sporadic resistance to cephalosporins has been reported. The third-generation fluoroquinolone, gatifloxacin, is highly effective against all known clinical strains of *S. typhi* both *in vitro* and *in vivo* and to date gatifloxacin-resistant strains have not been reported, although they are likely to be encountered in the future.

The guidelines of the Indian Association of Paediatrics (IAP) issued in 2006 recommend cefixime as the initial best-guess therapy for uncomplicated typhoid fever with azithromycin as the second-line antibiotic. For complicated typhoid fever, IAP recommends ceftriaxone with aztreonam and imipenem as second-line agents for complicated cases. In sub-Saharan Africa and the Middle East, the rate of resistance to fluoroquinolones remains relatively low and thus, in these regions, the initial treatment is with oral ciprofloxacin for uncomplicated disease and intravenous ciprofloxacin for severe or complicated infections. In randomized controlled trial, dexamethasone has been shown to decrease mortality (from 55% to 10%) in severe typhoid fever cases complicated by delirium, obtundation, stupor, coma or shock.

Surgical treatment

Surgical treatment is reserved for complications which most commonly include ileal perforation with peritonitis, intra-abdominal abscesses and much less commonly gangrenous typhoid cholecystitis. The surgery is undertaken under intravenous anti-*S. typhi* antibiotic cover and includes thorough cleansing of the peritoneal cavity and suture of the perforation with drainage. The edges of the perforation are trimmed before closure. Small bowel resection is indicated only for patients with multiple ileal perforations. If antibiotic treatment fails to eradicate the hepatobiliary carriage, then a cholecystectomy is indicated; however cholecystectomy is not always successful in eradicating the carrier state because of persisting hepatic infection.

Gastrointestinal (abdominal) tuberculosis

Worldwide, TB causes 3 million deaths annually and, approximately, one-third of the world's population is infected with the TB bacillus, although many such individuals do not develop clinical disease unless the immune system becomes depressed by disease such as HIV infection or immunosuppressive drugs and old age. Gastrointestinal TB has become a major health problem in both developed and underdeveloped countries. The recent significant increased incidence in Western countries has been the consequence of the HIV pandemic and increased immigration from countries with high prevalence of TB. Another important change concerns the reduced association between pulmonary and intestinal disease, such that only a minority of patients (<20–25%) with abdominal TB have abnormal chest radiographs indicative of active or healed pulmonary disease. Although gastrointestinal TB may affect any part of the gastrointestinal tract, including the oesophagus, the majority of infections involve the ileum and colon. The diagnosis of abdominal TB often proves difficult. In practice it is based on the demonstration of acid-fast bacilli, caseating or non-caseating epithelioid cell granulomas on histology, evidence of TB at other sites, and, in some cases, on complete response to antituberculous therapy. Multidrug-resistant TB is an increasing problem and, when present, more than doubles the 1 year mortality (from 20% to 50%).

Pathology

The disease is caused by *Mycobacterium tuberculosis*. The routes of gastrointestinal infection are (1) ingestion of infected sputum by patients with active pulmonary TB, especially those with pulmonary cavitation and positive sputum smears; (2) spread by haematogenous route from pulmonary tuberculous to mesenteric and submucosal lymph nodes; and (3) local spread from intra-abdominal organs involved by primary tuberculous infection (renal TB).

Pathologically gastrointestinal TB is characterized by inflammation and fibrosis of the bowel wall and the regional lymph nodes. Mucosal ulceration results from necrosis of Peyer's patches and from vascular thrombosis. At this stage of the disease, the pathological changes are reversible with antituberculous therapy and healing without scarring is possible. However, with progress of the disease, the ulcers become confluent and the continued tissue damage leads to extensive fibrosis with thickening of the bowel wall and the formation of pseudotumoral mass lesions. Strictures and fistula formation may also occur. The serosal surface of the bowel may exhibit nodular masses or miliary tubercles. The mucosa is inflamed with hyperaemia and oedema similar to that encountered in Crohn's disease but short of the classical cobblestone appearance. Aphthous ulcers may develop in the colon in some patients. Caseation is not always present in the bowel wall but is almost universally present in the regional lymph nodes. Various types of granulomas are seen on histology: caseating, non-caseating, confluent, discrete and suppurative. Granulomas are most commonly found in the submucosal layer in association with a predominant lymphoplasmacytic inflammation in the lamina propria. Both intestinal perforation and bowel ischaemia may occur.

Intestinal TB can be classified into three macroscopic forms.

1. The ulcerative form (60%) – multiple superficial ulcers largely confined to the epithelial surface with their long axis lying perpendicular to the long axis of the bowel.
2. The hypertrophic form (10%) – causes thickening of the bowel wall with scarring, fibrosis, producing a firm matted mass which may simulate a carcinoma.
3. The ulcerohypertrophic form (30% of patients) – with features of both the ulcerative and hypertrophic forms.

The *hypertrophic type* affects predominantly the ileocaecal region and is characterized by the absence of gross caseation but with marked thickening of the submucosal and subserosal layers. It can also involve the ascending and transverse colon and is generally regarded as a low-virulence infection in a patient with a high degree of immunological resistance from previous exposure.

The *ulcerative type* affects predominantly the terminal ileum where multiple deep ulcers develop and extend to the serosa and may give rise to perforation. The serosal surface is thickened and studded with tubercles. Healing may result in multiple strictures with intervening dilated segments of ileum. Bacterial overgrowth may develop at this stage and cause diarrhoea and malabsorption. The *hypertrophic variety* is the least common and causes extensive matting of the small intestine and colon with considerable fibrosis and thickening.

Clinical features

Intestinal TB can occur in persons of any age, although it is more common in children and in older people. The ileum is more commonly involved than the jejunum and ileocaecal involvement is encountered in 80–90% of patients, presumably because of the abundant lymphoid tissue (Peyer's patches). Proximal (jejunal) small intestinal disease is seen more commonly with *M. avium intracellulare* complex infection. Intestinal obstruction when it occurs may be partial or complete. Segmental involvement causes the stenotic form of the disease. A recent report identified type II diabetes mellitus (23%) and alcoholism (23%) as the main comorbid disorders associated with abdominal TB.

Non-specific symptoms such as weight loss and abdominal pain are present in 80–90% of patients. Approximately 30% of patients have constipation. Nausea and vomiting may indicate the presence of subacute intestinal obstruction. The more specific clinical features of abdominal TB include abdominal pain, weight loss, anaemia and fever with night sweats.

Others present with a tender palpable mass in the right iliac fossa. Some patients may present acutely with symptoms of intestinal obstruction or acute right iliac fossa pain. Other acute presentations include haemorrhage and perforation, although free perforation is rare. Malabsorption is common and may be caused by subacute obstruction leading to bacterial overgrowth, but may be due to other causes including involvement of the mesenteric lymphatic system (tabes mesenterica) which may impair chylomicron transport resulting in steatorrhoea.

The physical signs include reduced muscle mass, abdominal tenderness, ascites, abdominal gaseous distension (subacute obstruction), abdominal mass (usually in the right iliac fossa) but may be felt elsewhere (omental cakes).

Diagnosis

The diagnosis of abdominal TB often poses difficulties even in endemic areas. This is documented by postmortem reports of patients in whom abdominal TB is diagnosed only after death. In essence the diagnosis of abdominal TB can only be firmly established by demonstrating acid-fast bacilli on smear or positive mycobacterial culture from the tissue or by demonstrating caseating granulomas by histological examination. The problem is, however, that abdominal TB is paucibacillary. In practice, the diagnosis requires a high index of suspicion in all cases together with a battery of laboratory, haematological, imaging and endoscopy tests. Even so, cases are encountered when despite strong clinical suspicion all the available tests are inconclusive. In these situations, provided malignancy can be excluded with reasonable certainty, a therapeutic trial of antituberculous therapy may be justified. However, laparotomy is indicated where malignancy cannot be ruled out. In these cases, a frozen section is recommended and a mesenteric lymph node should be removed for both culture and histology (caseating granulomas). Likewise, ascitic fluid should be sampled for staining for acid-fast bacilli and culture/guinea pig inoculation.

Attempts should be made to culture the mycobacterium from gastric washings, faeces, peritoneal fluid and tissue biopsies including enlarged peripheral lymph nodes. It is important to realize, however, that certain atypical mycobacteria (not responsible for TB) may be demonstrated in certain chronic inflammatory conditions including Crohn's disease. Haematological tests show non-specific changes of anaemia, hypoalbuminaemia and elevated ESR, but these are unhelpful in establishing the diagnosis. Likewise, the tuberculin test even when positive is unhelpful as it does not differentiate between an active and inactive disease. The serological tests: soluble antigen fluorescent antibody (SAFA) and enzyme-linked immunosorbent assay (ELISA) lack sensitivity and specificity. At best, they only suggest a probable diagnosis.

A recently developed RD-1 gene-based assay (the PBMC ELISPOT assay) for diagnosing TB infection shows promising results with a reported sensitivity and specificity of 89% and 78%, respectively. In patients with tuberculous ascites, sampling of the peritoneal fluid may be helpful if it contains $>30\,g/L$ protein and $>1000/mm^3$ (mostly lymphocytes), an ascitic/blood glucose ratio of <0.96, and adenosine deaminase levels of $>33\,IU/mL$. Acid-fast bacilli are rarely seen on smears but may be cultured from the ascitic fluid, especially after concentration by centrifugation prior to culture.

Imaging tests

- *Plain radiograph of the abdomen* is useful in patients with intestinal obstruction and perforation. It may also show calcification in lymph nodes. A chest radiograph is needed in all patients for evidence of active or healed pulmonary infection.
- *Barium contrast studies (follow-through or small bowel enema)* are very useful in suspected gastrointestinal TB as they can demonstrate bowel lesions such as multiple strictures. The other radiological findings identified by double contrast studies are mucosal irregularity, flocculation and fragmentation of barium (malabsorption), dilated, displaced (enlarged lymph nodes) and adherent fixed small bowel loops. Thickening of the ileocaecal valve lips and/or a wide gaping valve with narrowing of the terminal ileum (Fleischner sign) is said to be characteristic of TB. The mucosal ulcers seen with double contrast studies are initially shallow with elevated margins but become confluent as the disease progresses. The ulcers are typically perpendicular to the long axis of the bowel and their healing causes annular strictures. Although these findings are similar to those encountered in Crohn's disease, the cobblestone appearance is not encountered in intestinal TB. Ileocaecal TB is characterized by a deformed, narrow, irregular and short upwardly displaced caecum, with an incompetent ileocaecal valve, dilated ileum, an increased ileocaecal angle and shortened ascending colon.
- *Ultrasonography* is useful in demonstrating the presence of even small quantities of ascitic fluid and may detect peritoneal disease. The reported ultrasound findings include multiple, thin complete and incomplete septae and visible echogenic debris (high fibrin content of the ascitic fluid). Peritoneal thickening and nodularity are the other sonographic features of abdominal TB. Omental cakes and adhesions can also be detected but not peritoneal tubercles. Lymphadenopathy is seen as conglomerate masses and/or as scattered enlarged nodes with hypoechoic or anechoic centres from necrosis. Thickening of the SBM $\leq 15\,mm$ and increased mesenteric echogenicity combined with mesenteric lymphadenopathy are recognized characteristic sonographic features of early abdominal TB. Ultrasonography may also be useful for guiding ascitic fluid sampling and fine-needle aspiration cytology of lymph nodes/hypertrophic lesions.
- *Computed tomography:* the common findings on CT suggestive of abdominal TB are high-density ascites, lymphadenopathy, bowel wall thickening and irregular soft-tissue densities in the omental region. Abdominal lymphadenopathy is the commonest manifestation of abdominal TB on CT scanning. The lymph nodes involved most commonly include mesenteric, periportal, peripancreatic and upper para-aortic groups of nodes. Tuberculous lymphadenopathy commonly shows peripheral rim enhancement, frequently with a multilocular appearance, as distinct from lymphomatous adenopathy which typically exhibits homogeneous attenuation.
- The CT features that can help to distinguish tuberculous peritonitis from peritoneal carcinomatosis include a smooth peritoneum with minimal thickening and marked enhancement after contrast in TB as opposed to a nodular and irregular peritoneal thickening in peritoneal carcinomatosis.
- Comparative studies between CT and ultrasonography for the diagnosis of abdominal TB have shown that CT is more accurate than ultrasound

in detecting abnormalities such as periportal and peripancreatic lymph nodes and bowel wall thickening as, in contrast to ultrasound, there is no disturbance by bowel gas. However, bowel wall dilatation is better defined by ultrasound. Thus although CT is more sensitive and specific, ultrasound has the advantage of being less expensive, widely available, and easy to perform.

Endoscopy and laparoscopy

All types, upper gastrointestinal, colonoscopy, balloon enteroscopy and small bowel CE, have been used in the attempt to confirm the diagnosis. Endoscopic appearances in TB include hyperaemic nodular friable mucosa, irregular ulcers with sharply defined margins and pseudopolyps. It is important to stress that endoscopic biopsy which should always be performed may not reveal granulomas as these lesions occur most commonly in the submucosal layer although multiple biopsies from the same site may increase the yield. Endoscopic biopsy specimens may be subjected to polymerase chain reaction for detection of acud-fast bacilli.

Laparoscopy is invaluable and should always be undertaken in cases where the standard tests are inconclusive. The laparoscopic appearances of abdominal/peritoneal TB are virtually diagnostic: thickened peritoneum, multiple yellowish white/grey miliary tubercles studded over the parietal and visceral peritoneal surfaces. Some reports suggest that these appearances are more helpful in diagnosis of TB than either histological or bacteriological examination. Laparoscopy also enables procurement of tissue and fluid samples for laboratory and histological testing.

Treatment

All patients with abdominal TB are treated with a full course of antituberculous therapy using first-line drugs for 12–18 months, although some reported series have used shorter regimens lasting for 6–9 months with reported equally efficacy and increased compliance. Some recommend the addition of corticosteroids in patients with peritoneal disease in an attempt to reduce adhesive complications including sterility in females although there are no randomized controlled trials to confirm the efficacy of steroids.

Antituberculous drugs and regimens

The drugs are categorized as (1) *first-line drugs*, which are used in various combinations in the first instance; (2) *second-line drugs*, defined as such because they are less effective than the first-line drugs, or because they are more toxic or because they are unavailable in many developing countries; (3) *third-line drugs*, which are not on the WHO list either because they are not very effective or because their efficacy has not been proven with the exception of rifabutin (proven efficacy) but which is not included on the WHO list as it is considered too expensive for most developing countries.

All first-line antituberculous drugs have standard abbreviations:

- ethambutol is EMB or E
- isoniazid is INH or H
- pyrazinamide is PZA or Z
- rifampicin is RMP or R
- streptomycin is STM or S.

Drug regimens are also abbreviated: the drugs are listed using their single letter abbreviations. A *prefix* denotes the *number of months* of the treatment. A *subscript* denotes intermittent dosing (thus $_3$ = three times a week) and *no subscript* means daily dosing. Most regimens have an initial high-intensity phase, followed by a continuation phase (consolidation or eradication phase).

The regimens currently recommended are:

- 2HREZ/4HR = 2 months of izoniazid, rifampicin, ethambutol and pyrazinimide followed by 4 months of isoniazid and rifampicin
- 2SHRZ/4HR = 2 months of streptomycin, isoniazid, rifampicin and pyrazinamide followed by 4 months of isoniazid and rifampicin.

The WHO also recommends a 6 month continuation phase of HR (isoniazid and rifampicin) if the patient is still culture positive after 2 months of treatment.

The second-line drugs used in antituberculous therapy are:

- aminoglycosides: e.g. amikacin (AMK), kanamycin (KM)
- polypeptides: e.g. capreomycin, viomycin, enviomycin
- fluoroquinolones: e.g. ciprofloxacin (CIP), levofloxacin, moxifloxacin (MXF)
- thioamides: e.g. ethionamide, prothionamide
- cycloserine (the only antibiotic in its class)
- p-aminosalicylic acid (PAS or P).

The second-line drugs which are not included by the WHO are:

- rifabutin
- macrolides: e.g. clarithromycin (CLR)
- linezolid (LZD)
- thioacetazone (T)
- thioridazine
- arginine
- vitamin D
- R207910.

Other important measures include improving the general nutritional status of the patients by a high-calorie high-protein diet, physical rest and correction of deficiencies including hypoalbuminaemia and anaemia.

Surgical management

Although patients with intestinal obstruction usually require surgical treatment, successful resolution can occur although this often requires prolonged antituberculous therapy. Subacute intestinal obstruction or acute-on-chronic obstruction responds usually to conservative management and provides time for adequate investigation of the patient, who can then be managed electively if necessary. However, emergency surgical intervention is necessary for full blown intestinal obstruction with gross distension and for peritonitis due to small bowel perforation.

Ileocaecal resection with right hemicolectomy and primary anastomosis is now the standard operation for ileocaecal disease and the results are excellent provided that chemotherapy is maintained for appropriately long periods of time after surgery. For disease in regions other than the terminal ileum and caecum, segmental resection of the bowel with end-to-end anastomosis is performed. More recently stricturoplasty has been introduced to deal with fibrotic strictures. Again this must be performed under antituberculous drug therapy which has to be continued for some time after surgery. Other procedures which may be necessary include drainage of tubo-ovarian abscesses with or without salpingo-oophorectomy, gastrojejunostomy for gastric outlet obstruction caused by retroperitoneal and gastroduodenal lymphadenopathy and biopsies including harvest of mesenteric lymph nodes.

Actinomycosis

Actinomycosis is caused by filamentous, Gram-positive, non-acid-fast bacteria which belong to the order of Actinomycetales (family Actinomycetaceae, genus *Actinomyces*). However, other organisms, e.g. *Propionibacterium, Actinobaculum* and *Bifidobacterium* may cause similar disease. *Actinomyces* organisms are anaerobic/microaerophilic exhibiting slow growth to form colonies with a characteristic 'molar tooth' appearance. The most commonly isolated pathogens are *Actinomyces israelii, Actinomyces gerencseriae, Actinomyces turicensis, Actinomyces radingae* and *Actinomyces europaeus but several other species may cause the disease.* Nearly all actinomycotic lesions contain companion bacteria, the most important being *Actinobacillus actinomycetemcomitans, Peptostreptococcus, Bacteroides, Staphylococcus, Streptococcus* and Enterobacteriaceae, varying with the site of the disease. These companion pathogens enhance the low pathogenicity of actinomycetes. Actinomycosis is characterized by contiguous spread, suppurativa granulomatous inflammation, with the formation of multiple abscesses and sinus tracts which discharge 'sulphur granules'. The most common clinical forms of actinomycosis are cervicofacial, thoracic, abdominal and pelvic (in females).

Pathophysiology

Actinomycetes form part of the normal flora of the oral cavity, the lower gastrointestinal and the female genital tracts. The micro-organisms, because of their low virulence, require a break in the integrity of the mucous membranes and/or the presence of devitalized tissue to cause human disease. Actinomycosis is essentially a polymicrobial infection as it involves the co-operation of other bacterial species known as 'companion' pathogens which elaborate toxins that inhibit host defences. They are responsible for the early manifestations of the disease and may contribute resistance to antibiotic therapies.

Once infection is established, the host mounts an intense inflammatory suppurative, granulomatous response which is destructive and produces intense fibrosis. Actinomycosis spreads contiguously through tissue planes with direct involvement of tissues and organs in the path of the infection. The host response results in the formation of multiple abscesses which on bursting lead to sinuses discharging granular pus – the sulphur granules. Although lymphatic spread is unusual, haematogenous dissemination to distant organs may occur in any stage of the disease.

Cervicofacial actinomycosis

This is the most common type and accounts for 50–70% of reported cases. The infection typically occurs following oral extraction/surgery or in patients with poor dental hygiene. Initially, cervicofacial actinomycosis forms soft-tissue swelling of the perimandibular region. As the disease progresses, direct spread to the adjacent tissues occurs with the development of fistulas/sinuses discharging the typical yellowish pus containing sulphur granules. Direct invasion of the skull and bloodstream may occur in untreated cases.

Thoracic actinomycosis

Thoracic actinomycosis is less common (15–20% of reported cases) and usually arises following inhalation of infected oropharyngeal secretions. Rarely, thoracic actinomycosis results from oesophageal perforation or from direct spread from the neck or abdomen, or via haematogenous spread. Thoracic actinomycosis forms a pulmonary infiltrate or mass, which can invade the pleura, pericardium and chest wall, ultimately leading to external pleurocutaneous fistulas discharging sulphur granules.

Actinomycosis of the abdomen and pelvis

Actinomycosis of the abdomen and pelvis accounts for 10–20% of reported cases. Typically, these patients have a history of emergency bowel surgery, e.g. perforated acute appendicitis, perforated colonic diverticulitis or perforation by ingested foreign bodies (chicken or fish bones). The most common site of the disease is the ileocaecal region with presentation as a slowly growing mass in the right iliac fossa which may simulate malignancy. Involvement of any abdominal organ and the abdominal wall occurs by direct spread, with eventual formation of peritoneocutaneous fistulas discharging pus. Pelvic actinomycosis most often results from infection of the uterus occurring in association with longstanding intrauterine contraceptive devices.

Clinical features

Actinomycosis occurs worldwide, but with higher prevalence rates in areas with low socioeconomic status and in individuals with poor dental hygiene. The disease can affect all ages, but most reported cases are in young to middle-aged adults. The clinical features of cervicofacial actinomycosis include history of dental extraction or trauma to the mouth in an individual with poor oral hygiene, dental caries, or periodontal disease; painless or less commonly painful soft-tissue swelling of the perimandibular region and fever. The nodular lesion(s) due to abscess formation are usually located at the angle of the jaw. These gradually increase in size and number, burst to form sinuses/fistulas which open and discharge pus through either the cheek or submandibular region. The nodules are tender initially but become non-tender on palpation as the diseased tissues harden from the dense fibrosis. The skin overlying the diseased tissues is red with bluish discoloration. The patient experiences difficult and painful chewing of food from involvement of muscles involved in mastication and this may progress to trismus.

Patients with thoracic actinomycosis, in addition to poor oral hygiene, are individuals at risk of inhalation, e.g. epileptics and chronic alcoholics. The established disease causes chest pain, a dry or productive cough and shortness of breath with fever, weight loss, fatigue and anorexia. Aside from fever, the other signs include cachexia, abnormal breath sounds and haemoptysis and the development of pleurocutaneous fistulas discharging the typical granular yellow pus.

In abdominal actinomycosis, there is nearly always a history of emergency abdominal operation. Examination reveals scar(s) from previous abdominal surgery, low-grade fever and cachexia together with a right lower quadrant mass which is firm-to-hard in consistency, non-tender and fixed to underlying tissues. Infection spreads mainly to the psoas muscle, abdominal wall and adjacent organs. Spread of the infection by the portal venous system results in multiple intercommunicating loculated abscesses (honeycomb liver). The clinical features also include low-grade fever, weight loss, fatigue, altered bowel habits, abdominal pain, nausea and vomiting, and a tender mass usually in the right iliac fossa with varying induration of the overlying skin and the eventual development of the characteristic discharging sinuses. Peritoneocutaneous fistulas may discharge pus on the anterior abdominal wall or the perianal region. In females with pelvic disease, the clinical features include lower abdominal discomfort, abnormal vaginal bleeding (meno/metrorrhagia) or discharge and pelvic mass.

Treatment

The majority of cases require only antibiotic therapy, although surgery may be required in selected cases: drainage of abscesses, excision of sinus tracts and dense fibrotic lesions, decompression of closed-space infections, and for the relief of obstruction most commonly to the ureter.

Penicillin G is the drug of choice. High-dose penicillin is administered over a period of 6–12 months. To a large extent, the duration of antibiotic therapy depends on the stage and extent of the disease. Thus shorter courses (<6 months) are used especially in cases of early cervicofacial actinomycosis with minimal disease. Thus the period of antibiotic therapy is tailored to the individual patient and is dictated by the clinical and radiographic response. The development of resistance to penicillin by actinomycetes is not a problem and lack of a clinical response to penicillin usually indicates resistance by the companion pathogens. This is the usual reason for change in the antibiotic regimen. Another frequent problem is allergy to penicillin which precludes its use.

The antibiotic regimens used in the absence of allergy to penicillin are:

- penicillin G: adult dose 12–24 million units daily intravenously by continuous infusion or in 4 hourly divided doses for 1–2 weeks, followed by oral penicillin VK for 6–12 months
- penicillin VK: adult dose 500 mg orally every 6 hours for 6–12 months.

In patients with penicillin allergy, the following antibiotics are used:

- doxycycline (Vibramycin): adult dose: 100 mg orally or intravenously every 12 hours
- clindamycin but this does not cover the companion pathogens. Adult dose: 600 mg intravenously every 8 hours

In patients with poor response due to resistance by copathogens, the following antibiotics may be used:

- amoxicillin/clavulanic acid (Augmentin): adult dose 500 mg orally every 8 hours or 875 mg orally every 12 hours
- ceftriaxone: adult dose 2 g intravenously or intramuscularly every 12–24 hours
- imipenem/cilastatin – useful in moderately severe to severe forms of abdominal and pelvic actinomycosis: adult dose 500–1000 mg intravenously every 8 hours.

HIV enteropathy

Gastrointestinal problems are extremely common in patients with AIDS (Box 29.2). In some of these patients, the symptoms are caused by opportunistic infection with protozoal organisms, various bacterial species, viruses and fungi or tumours.

However, in a significant percentage of patients (about 30%) no identifiable pathogens can be isolated from the stool. This indicates that HIV may cause a specific enteropathy and recent studies using DNA probes have identified the presence of HIV in the base of the crypts and in the lamina propria. HIV enteropathy is nowadays defined as a syndrome of chronic (more than 4 weeks' duration) severe watery diarrhoea occurring in HIV-positive patients without an identified infectious cause despite intensive investigation. This 'idiopathic' enteropathy is thought to be due to direct effects of the HIV on the enteric mucosa. The intestinal mucosal immune system is now known to be involved in the pathogenesis of AIDS, with key events of the disease (transmission, viral amplication, and destruction of CD4+ T-cells) taking place within the gastrointestinal tract.

Histologically, the enteropathy is characterized by lymphocytic infiltration of the mucosa with evidence of damage, including villous atrophy, crypt hyperplasia and villous blunting of the epithelial layer. These changes appear to result from a direct

BOX 29.2 Aetiology of diarrhoea in patients with AIDS

- Protozoal
  - *Cryptosporidia*
  - *Isospora belli*
  - *Giardia lamblia*
  - *Entamoeba histolytica*
- Bacterial
  - *Salmonella*
  - *Shigella*
  - *Yersinia*
  - *Mycobacterium avium intracellulare*
  - *Campylobacter* species
  - *Clostridium difficile*
- Viral
  - Cytomegalovirus
  - Rotavirus
  - Norwalk virus
- Fungal
  - *Candida albicans*
  - *Blastocystic hominis*
- Neoplasm
  - Lymphoma
  - Kaposi's sarcoma
- Idiopathic
- AIDS enteropathy

toxic effect of the HIV on the enterocytes mediated by the HIV accessory protein *Tat* which inhibits glucose uptake by the enterocyte. Another HIV protein *gp120* decreases the ability of the gut mucosal cells to maintain ionic gradients. There is, in addition, activation of the immune system with high levels of proinflammatory mediators (IL-6, IL-10 and interferon-γ). The most significant change consists of the substantial loss of the CD4+ T cells and counts below 200 cells/μL are invariably associated with HIV enteropathy.

Clinical features

HIV enteropathy can develop in the acute phase of AIDS or present in patients with advanced disease. The enteropathy is characterized by severe watery diarrhoea associated with increased inflammation and gut permeability. Other gastrointestinal symptoms include weight loss, abdominal pain, fever, sore throat, lymphadenopathy, arthralgia and an erythematous, maculopapular rash.

To some extent, the clinical presentation of HIV-associated diarrhoea varies with the principal site of involvement of the gut. Thus small bowel diarrhoea tends to result in large bulky postprandial stools immediately after eating, and the individual may experience postprandial paraumbilical abdominal pain and weight loss. Moreover, the diarrhoea abates significantly with fasting. In contrast patients with large intestine diarrhoea (colitic diarrhoea) have frequent, small-volume stools, often containing visible blood and mucus. These individuals often complain of lower quadrant abdominal pain, and rectal urgency/tenesmus. However, mixed clinical pictures are common and it is often difficult to differentiate between small and large bowel disease.

Diagnosis

In AIDS patients who develop symptoms of enteropathy, the initial step is three consecutive stool cultures for *Salmonella*, *Shigella flexneri* and *C. jejuni*, in addition to assays for *Clostridium difficile* toxins. This is followed by upper gastrointestinal endoscopy and colonoscopy for inspection of the mucosa and for procurement of biopsy material. Biopsies are cultured for cytomegalovirus and mycobacteria, adenovirus and herpes simplex virus. Duodenal fluid samples are also examined for parasites.

The routine diagnosis of AIDS itself is based on the detection of antibodies to HIV. Following infection, antibodies to the various antigens associated with the virus often do not appear for about 6 weeks, so that there is a long window when an individual may be negative on antibody testing to HIV and yet still be infected. Other tests include detection of circulating viral p24 antigen and the use of polymerase chain reaction by which very small amounts of viral RNA in infected cells can be greatly amplified.

Treatment

If a specific infection is established treatment is directed to eradicating the organism responsible. Symptomatic relief can be obtained using non-specific antidiarrhoeal therapy. Patients will benefit from supportive therapy with rehydration and electrolyte supplements. In severe cases, total parenteral nutrition is indicated. Plasma or serum citrulline assays have been introduced as an accurate assessment of the enterocyte cell mass. Citrulline is a metabolite of glutamine and is synthesized by small bowel enterocytes. Citrulline plasma levels are not influenced by nutritional status, level of hypoalbuminaemia or inflammatory status but are elevated in the presence of significant renal failure (creatinine clearance of <30 mL/min) because citrulline is normally metabolized to arginine in the proximal convoluted tubules. In the absence of renal impairment, plasma citrulline assays thus provide a reliable measure of the disease severity in HIV enteropathy and of the need for parenteral nutrition which is indicated in patients with a citrulline level of <10 μmol/L.

Antiretroviral therapy (ART) reduces plasma viral loads with consequent increase in peripheral blood CD4+ T-cells. In HIV enteropathy ART is effective in decreasing the gastrointestinal symptoms. However, a decrease in viral replication and CD4+ T-cell reconstitution is seldom obtained with standard ART in small bowel HIV enteropathy, and a much better response in these patients is obtained with highly active antiretroviral therapy which results in much greater reconstitution (twofold) of CD4+ T-cells.

Parasitic infestations of the small bowel

Ascaris (roundworm)

This infestation occurs directly from person to person without an intermediate host. The responsible parasite is *Ascaris lumbricoides* which enters by the oral route. Ova which have a special lipid layer which resists acid digestion in the stomach remain viable in the soil for several months to years. Immediately after ingestion the larvae emerge from the egg and penetrate the duodenal wall to enter the bloodstream to reach the liver, heart and pulmonary circulation where they enter the pulmonary alveoli at approximately 3 weeks from ingestion. They are then coughed up to be swallowed into the gut where they mature as adult worms (up to 40 cm in length) in the jejunal lumen. Infestation with *A. lumbricoides* is endemic throughout the world but the highest rates are encountered in China, Southeast Asia, in coastal regions of West Africa and in central Africa. Risk factors include low socioeconomic class and poor sanitation.

Clinical features

Initially, the symptoms are caused by the larval migration especially to the lungs (wheezing, fever, cough, chest pain and dyspnoea) including established asthmatic attacks. Migration to other sites may cause other symptoms, e.g. skin rash, conjunctivitis and convulsions.

The symptoms caused by the adult worms include abdominal pain with colicky attacks, nausea, vomiting, anorexia, weight loss, diarrhoea, malabsorption and anal itching. However, many infested individuals remain asymptomatic although they are at risk from complications which include intestinal distension and obstruction with vomiting (sometime of worms) and colicky abdominal pain, intestinal perforation with peritonitis, appendicitis, acute pancreatitis, volvulus, cholangitis and liver abscess.

Diagnosis

The diagnosis is made by demonstrating ova in a faecal sample, although adult worms are sometimes passed per rectum the worms may be seen on a small bowel follow-through. A full blood count may reveal eosinophilia and usually anaemia. Hypoalbuminaemia is frequent and the liver function tests may reveal evidence of liver damage. Other tests may be needed depending on the clinical picture including endoscopic retrograde cholangiopancreatography (ERCP) if involvement of the biliopancreatic tree is suspected.

Treatment

Eradication treatment is usually with one of the benzimidazole compounds such as mebendazole or albendazole or piperazine. Surgery is indicated for intestinal obstruction which does not resolve on conservative measures. At operation an attempt should be made to manipulate the mass of tangled worms into the caecum. If this fails, it may be necessary to extract the worm bolus via an enterotomy. Occasionally, a resection may be necessary if the viability of the bowel is in any doubt or if a perforation has occurred.

Mebendazole has a cure rate of 76–95% and is used as the first-line treatment of choice for adults and children over the age of 2. In some regions where mebendazole has been used for long periods, resistance to the drug has been reported. Piperazine is used in children (between 3 months and 2 years) but can also be used as an alternative to mebendazole in adults usually combined with a purgative (senna) to ensure complete expulsion. This regimen has been reported to be successful in relieving intestinal obstruction, thereby avoiding an operation, and can be instilled in the bile duct during ERCP in patients with biliary ascariasis.

Levamisole (nicotinic acid antagonist) is highly effective and is given as a single dose.

Ancylostomiasis (hookworm)

Ancylostomiasis is endemic in most tropical countries. It is caused by two filarial worms, *Ancylostoma duodenale* and *Necator americanus*. The larvae enter the body by penetrating into skin, usually the sole of the foot, forming an itchy lesion. The larvae enter the bloodstream to reach the lungs and through the alveoli, the tracheobronchial tree from which they ascend to the laryngopharynx where they are swallowed. *A. duodenale* can also be acquired by eating the raw or undercooked flesh of infected animals as larvae can penetrate the skin of various animals and survive in their muscle tissue.

The adult worm lives attached by hooks to the mucosa of the jejunum and ova are excreted in the faeces. Larvae develop from the eggs deposited on the soil in about 1 week. Infection is contracted by treading on the larvae during bare foot walking. The warm humid environment in mines enables the survival of hookworm larvae for long periods and accounts for the high prevalence of ancylostomiasis in miners in whom it is sometimes referred to as 'miners' anaemia.

Clinical features

Infestation may be associated with severe dyspepsia and gastrointestinal bleeding causing a microcytic anaemia and frank gastrointestinal haemorrhage especially in children. However, ancylostomiasis can produce minimal symptoms if the infestation is mild and indeed the disease may be asymptomatic if only a few worms are present. Symptoms are usually present in patients with longstanding infestations with high worm loads. They include asthenia, tiredness, weight loss and weakness in association with iron-deficiency anaemia. More severe symptoms, e.g. abdominal pain, overt malnutrition, bloody diarrhoea and heart failure, are well documented. To some extent, the range and severity of symptoms vary widely and depend on the nutritional status and overall general health of the patient, the species of worm and the extent of infestation.

Diagnosis

The diagnosis is made by the identification of ova in the faeces or adult worms in duodenal or jejunal fluid. A peripheral blood eosinophilia may occur during the invasive stage.

Treatment

The first stage of treatment is the correction of anaemia using either oral or injectable iron. Benzimidazole compounds, mebendazole or albendazole or pyrantel pamoate are used to clear the infestation.

Small bowel conditions causing malabsorption

Small bowel bacterial overgrowth

In this syndrome, also known as dysbiosis, the small intestine becomes colonized by bacteria with an increase in the concentration of organisms which are usually confined to the lower small bowel and the colon. This is usually caused by surgery or disease which results in excess bacteria entering the small intestine, or from delayed clearance of bacteria due to stasis (stagnant or blind loop syndrome). The various causes of bacterial overgrowth include:

- excessive entry of bacteria into the small intestine
  - achlorhydria – absence of bactericidal gastric HCl
  - gastrojejunostomy
  - partial/total gastrectomy
  - entercolic fistulas
  - cholangitis
  - loss of ileocaecal valve following right hemicolectomy
- intestinal stasis
  - Crohn's disease – stenosis
  - TB – stenosis
  - small bowel diverticulosis – stasis
- afferent loop stasis
  - enteroenteric anastomosis and other intestinal bypass procedures
  - subacute obstruction – adhesions, strictured anastomosis
  - blind loops
  - diabetes mellitus – autonomic neuropathy
  - radiation eneteritis – stenosis, impaired intestinal motility
  - scleroderma – impaired intestinal motility
  - amyloidosis.

In some cases bacterial overgrowth may develop in the absence of an obvious local cause particularly in patients with malnutrition or immune deficiency. There is also a growing body of research linking small bowel bacterial overgrowth (SBBO) with chronic fatigue syndrome, fibromyalgia and irritable bowel syndrome.

Pathophysiology

The bacterial overgrowth in the small intestine is usually in the order of 10^7 to 10^9 CFU/mL of intestinal contents. Bacterial colonization results in intestinal mucosal injury characterized by patchy inflammatory changes in the lamina propria which are accompanied by alterations in the concentrations of the brush border enzymes. The increased bacterial population deconjugates the intraluminal bile salts by removing the glycine or taurine moiety. In addition the bacteria dehydroxylate the steroid nucleus at the C7 position with the formation of deoxycholate and lithocholate which have a tendency to precipitate at the intraluminal pH levels of the small intestine. Thus these bile salts participate poorly in the emulsification of fat and they are passively absorbed to a limited extent by non-ionic diffusion through the small intestine. This leads to a net reduction of the concentration of effective bile salts in the lumen of the small bowel to levels below that required to form micelles. This in turn results in malabsorption of fats and fat-soluble vitamins.

The bacterial species which are responsible for bile salt deconjugation are eubacteria, bacteroides and corynebacteria. These bacteria also bind vitamin B_{12} and convert it to inactive derivatives (cobamides) which block the ileal receptors for the vitamin as well as for intrinsic factor. The resulting malabsorption of vitamin B_{12} may lead to megaloblastic anaemia. Often, however, the anaemia has a dimorphic picture because of an iron-deficiency component due to chronic blood loss from the primary lesion itself or from malabsorption of iron. Folate deficiency is rare as the bacteria synthesize folate in substantial amounts and some patients may in fact exhibit a high serum folate.

There is some malabsorption of carbohydrates and proteins although this is rarely significant. The main reason for malnutrition in patients with bacterial overgrowth is a diminished dietary intake, which accounts for growth retardation in children. The bacteria also metabolize triglycerides to free fatty acids which they hydoxylate to form hydroxy fatty acids. These impair the absorption of water in sodium by the intestinal mucosa of both the small and large intestine by acting as laxatives. This together with the action of the dehydroxylated bile salts, enterotoxin and osmotic load created by the fermentation of the major dietary components accounts for the diarrhoea which is seen in these patients.

Clinical features

The symptoms of bacterial overgrowth are varied. Initially the patients may exhibit symptoms referable to the underlying pathology, e.g. postgastrectomy symptoms or recurrent intestinal colic due to subacute obstructing lesions. Patients with jejunal diverticulosis (Figure 29.30) are symptomless until the onset of malabsorption, although they may occasionally develop unrelated complications such as gastrointestinal bleeding or perforation of one of the diverticula leading to generalized peritonitis. The symptoms and signs of bacterial overgrowth itself are often non-specific and include malaise, nausea and vomiting, excessive borborygmi and weight loss. Diarrhoea is extremely common and is usually watery but frank steatorrhoea with bulky pale offensive stools which are difficult to flush is less common.

Diagnosis

A full blood count is essential to assess the extent and type of anaemia which is often present in these patients. Estimations of urea and electrolytes, albumen and total protein levels are also important. The most useful test for assessing the presence of bacterial overgrowth, small bowel however, is the hydrogen breath test (see section small bowel investigations).

Complications

The complications of bacterial overgrowth include glossitis, stomatitis, anaemia, hypoproteinaemia with peripheral oedema, tetany, osteomalacia and rickets and growth retardation in children. Occasionally neurological manifestations such as paraesthesiae and peripheral neuropathy may be found in association with vitamin B_{12} deficiency.

Treatment

Surgical treatment of the underlying conditions wherever possible is the definitive and curative treatment. However, situations are frequently encountered where surgical treatment is

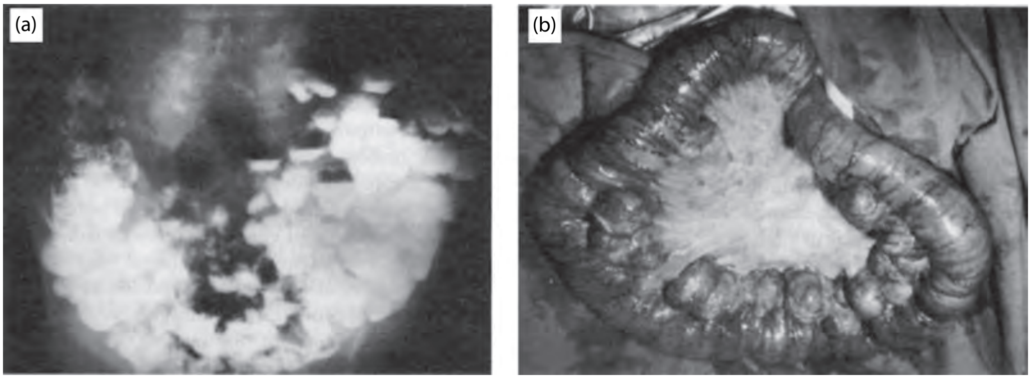


Figure 29.30 (a) Small bowel barium study showing extensive and multiple small bowel diverticula in a 65 year old woman with severe malabsorption. The steatorrhoea improved considerably with oral tetracycline therapy. (b) Multiple jejunal diverticula localized to a segment of the upper small intestine. The patient's bacterial overgrowth syndrome resolved after excision of the affected bowel.

not possible or advisable because of the extensive or systemic nature of the underlying disease, e.g. extensive jejunal diverticulosis and scleroderma. In these cases intermittent therapy with oral antibiotics is often beneficial. As control of the anaerobic organisms seems to be the most helpful objective, metronidazole is generally preferred, although tetracycline is also of value. A course of antibiotics is usually administered for about 2 weeks at a time and although repopulation of the intestine by bacteria occurs soon after discontinuation of the treatment, symptomatic improvement may last for several months. Intermittent therapy is therefore preferred to long-term management.

The treatment also includes dietary changes and the use of probiotics/prebiotics. The diet recommended in the treatment of SBBO is free of simple sugars and grains/cereals and low in fruit and starchy vegetables. This is thought to work by restricting the nutrition available for bacteria in the upper gastrointestinal tract and reduces the amounts of alcohol and organic acids produced by bacterial fermentation.

Specific therapy is with antibiotic medication together with natural antibiotic substances. Therapy usually requires several months to achieve control or reversal of the bacterial overgrowth of the small intestine. The antibiotics used are metronidazole, tetracycline and ciprofloxacin. In many ways all three have the potential to both cause and treat SBBO. The antibiotic used should target only the pathogenic bacteria without suppressing the beneficial flora. This is, however, more easily said than done. Metronidazole is the antibiotic of choice in most cases of SBBO. It is extremely effective against anaerobic bacteria but has little effect on aerobes. The most common side effects of metronidazole include nausea, diarrhoea and abdominal discomfort in addition to vaginal candidiasis in females. Thus metronidazole therapy is best accompanied by use of probiotics and of possibly antifungal agents.

Tetracycline is sometimes used when therapy with metronidazole is not successful. It is especially important to reinoculate the gut with probiotics when tetracycline is used. Ciprofloxacin is the most common drug used when the overgrowth is by aerobic bacteria. The side effects of this antibiotic include nausea, diarrhoea, abnormal liver function tests, vomiting and rash.

Probiotic bacteria

The most numerous probiotic bacteria present normally in the small bowel are species of lactobacilli (*Lactobacillus acidophilus*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, etc.), whereas bifidobacteria (*Bifidobacterium bifidum*, *Bifidobacterium longum*, *Bifidobacterium infantis*, etc.) are the probiotics that predominate in the normal colon. Most of the probiotic products consist of one or more species of bacteria from these two groups. Lactobacilli impart considerable benefit to the gut by preventing overgrowth by pathogens, aid in the digestion of lactose and dairy products, improve nutrient absorption and maintain the integrity of the intestinal epithelium preventing access of macromolecules into the bloodstream.

Bifidobacteria also prevent colonization by pathogenic bacteria, maintain the low pH balance by production of acids, inhibit the growth of nitrate-forming bacteria, synthesize B-complex vitamins and help to regulate peristalsis.

Prebiotics

These are ingestible oligosaccharides which provide food and thus promote the growth of probiotic bacteria. They are present in certain fruits and vegetables, e.g. bananas, asparagus, garlic, onions, wheat, tomatoes, onions, etc. They are also available as supplements: fructo-oligosaccharides, galacto-oligosaccharides and inulin.

Short gut syndrome

This is the most serious form of intestinal decompensation and is encountered after extensive resection of the small bowel.

Pathophysiology

The outcome following intestinal resection depends on (1) the extent and site of resection, (2) the age of the patient and (3) the physical and mental condition of the patient.

The intestinal decompensation that follows massive resection of the small intestine is due to a sudden reduction of the absorptive area and a greatly reduced transit time which further aggravates malabsorption and in extreme cases limits the extent of digestion. Malabsorption of fats and proteins is invariably present but its severity varies with the length and site of the residual bowel. Carbohydrate absorption is less severely affected and is the first to return to normal in 4–6 weeks. The improvement in the absorptive capacity for protein is more gradual and the absorption of fats remains impaired at a fixed percentage. This is the result of interruption of the enterohepatic circulation of bile salts after ileal resection and a diminished bile salt pool, as the increased hepatic synthesis is unable to compensate for the large daily faecal losses.

Spillover into the colon of primary bile salt conjugates where bacterial action will convert them into deconjugated and dehydroxylated derivatives contributes to the diarrhoea which is characteristic of the short gut syndrome. This is because some of these secondary bile salts block the absorption of water and electrolytes by the colonic mucosa. The hydroxylation of the unabsorbed fatty acids by colonic bacteria has also been implicated in diarrhoea. The malabsorption of fats and fat-soluble vitamins is accompanied by the malabsorption of calcium and magnesium which are precipitated as soaps with the unabsorbed fatty acid. The severe diarrhoea which can amount to several litres in the first few weeks after resection is further aggravated by the low pH of the stool and often causes severe perianal irritation.

Lactose intolerance may occur in some patients after extensive small bowel resection and this has been attributed to a rapid transit time allowing insufficient time for lactase to act and to a reduction of the total intestinal lactase activity.

Aetiology

The aetiology is essentially extensive resection of small bowel for any of the following reasons:

- Crohn's disease
- mesenteric infarction
- radiation enteritis

- midgut volvulus
- multiple fistulas
- small bowel tumours.

Crohn's disease is by far the commonest cause as this often necessitates repeated resections over a number of years culminating in the short gut syndrome. Hence the reason for a conservative approach in these patients with limited resections and the judicious use of stricturoplasty for short fibrous strictures. While resection of more than half of the small bowel is frequently accompanied by serious malabsorption there is debate as to the extent of the small bowel resection which results in the short gut syndrome. An assessment of the length of the residual small bowel after resection is essential as there is a crucial length below which intestinal decompensation of varying degrees takes place. Patients with a residual small bowel length of less than 2 m have a diminished work capacity and those with less than 100 cm require home parenteral nutrition on an indefinite basis. Infants and neonates seem to tolerate extensive small bowel resections better than adults and a minimum of 30 cm of small intestine can support both nutrition and growth.

The critical length of residual small bowel necessary to prevent serious maldigestion and malabsorption depends on the site of resection and the retention or otherwise of the ileocaecal valve. Thus ileal resections are less well tolerated than jejunal resections, largely because the active transport sites for bile salts and vitamin B_{12} are localized in the ileum. Whenever possible, the surgeon should retain a terminal segment of ileum together with the ileocaecal valve. An intact ileocaecal valve slows the transit time and limits the degree of colonization of the residual intestine by an overgrowth of colonic bacteria.

Clinical features

The main clinical features are weight loss and diarrhoea which is watery and frequently copious. Because of diarrhoea and the low pH of the stool, perianal irritation is common. In patients who develop lactose intolerance the consumption of fresh milk is followed by attacks of abdominal pain and an increase in the severity of the diarrhoea. Other clinical features depend on the development of complications (see below).

Diagnosis

Diagnosis is made on clinical grounds with information on the extent of small bowel resection. Other tests of malabsorption (see section Small bowel investigations) are helpful in developing an objective assessment of the severity of the malabsorption.

Course of condition

With time, structural and functional changes occur in the residual bowel as part of the process of adaptation to the reduced absorptive area. In the human, the structural changes consist of dilation of the remaining intestine and enlargement of the villi. An increase in the length of the bowel has also been reported in the neonate. These structural changes are accompanied by a gradual improvement in the absorption of water, electrolytes, carbohydrates and protein. The enhanced absorption of water-soluble substances per unit length of intestine is not the result of greater absorption by individual cells but rather from an increased cell population in the intestinal villi. Both luminal and humoral factors are involved in this adaptive hyperplasia. The luminal factors include alimentary secretions and ingested nutrients. Duodenal juice may be particularly important as a source of epidermal growth factor. The maintenance of adequate intraluminal nutrition by the oral ingestion of appropriate food is essential for the adaptive response. Enteroglucagon is the most important humoral agent influencing the adaptive response. The stimulus for its release is thought to result from increased exposure of the residual bowel to luminal nutrition. The enteroglucagon then acts on the mucosal cells and stimulates an increased turnover of intracellular polyamines and cell growth. There is evidence however that the effect of intraluminal nutrition is also mediated through direct contact of the nutrients with the epithelial cells. Other factors which may contribute to compensation by the residual intestine include the changes in the motility pattern leading to a gradual slowing of the transit time and increased absorption of water-soluble substances by the colon.

In cases of mesenteric infarction varying amounts of colon are frequently resected with the small intestine. It is not known whether the extent of the associated colonic resection influences the clinical outcome but colonic mucosa does absorb water and salt and to a lesser extent glucose and amino acids. Some reports indicate that residual colon minimizes the severity of diarrhoea and may assume some of the absorptive functions of the small intestine.

Complications

The complications which can be associated with short gut syndrome include:

- gastric hypersecretion
- cholesterol gall stones
- hepatic disease
- impaired renal function and stone formation
- metabolic bone disease.

Gastric hypersecretion

Gastric hypersecretion occurs transiently and bears no relation to the extent of the intestinal excision although it appears to be more common after proximal resections. It is caused by an increased rate of basal acid secretion consequent on delayed clearance of gastrin. Around 50% of patients with extensive small bowel resection are affected.

Gall stone formation

There is a well-recognized increase in the incidence of gall stones after extensive ileal resection and jejunoileal bypass for morbid obesity. Cholesterol stones are the result of a reduced bile salt pool which leads to the bile becoming lithogenic, i.e. supersaturated with cholesterol which precipitates as cholesterol crystals. In addition however, pigment gall stones also occur after ileal resection.

Hepatic disease

Mild hyperbilirubinaemia, elevation of the serum transaminases and impaired excretion of BSP often occur after both massive intestinal resections and jejunoileal bypass. These are associated with fatty infiltration of the liver with or without hepatic atrophy. The increased deposition of fat in the liver after jejunoileal bypass occurs in patients while they are actively losing weight but not when the weight is stationary. Acute fulminant liver failure has been reported most commonly after jejunoileal bypass. The onset of the liver failure usually occurs within the first 6 months after operation and the clinical symptoms are those of a flu-like illness with anorexia, nausea, vomiting, rapid weight loss and a fall in the serum albumen. The onset of these manifestations is an indication for restoration of intestinal continuity. Cirrhosis may develop after jejunoileal bypass. Initially it causes no symptoms and its onset can only be detected by serial liver biopsies.

Impaired renal function and stone formation

The severe diarrhoea may lead to fluid and electrolyte losses and to a decrease in the glomerular filtration rate with a tendency to reduce urine output and concentrated urine. The consequent loss of electrolyte-rich intestinal fluid causes hyponatraemia and hypokalaemia. A metabolic acidosis ensues from the fixed base losses into the gastrointestinal tract with the excretion of persistently acid urine. The urinary calcium excretion is low but the oxalate concentration and excretion is high, especially in patients with an intact colon. Urinary calculi of all types (urate and oxalate) are very common in patients after small bowel resection and the vast majority of patients will develop this complication during their lifetimes.

The water and salt depletion results in increased secretion of aldosterone in an attempt at maximal renal salt conservation but at the expense of increased potassium loss. This chronic hypokalaemia apart from causing muscle weakness, anorexia and cardiac arrhythmias may limit protein synthesis and impair transport and utilization of carbohydrates. In addition it may lead to the development of a renal tubular nephropathy consisting of a characteristic patchy vacuolar change in the cells of the proximal convoluted tubules. This syndrome of hypokalaemic vacuolar nephropathy can complicate other disorders associated with diarrhoea including ulcerative colitis and coeliac disease.

Metabolic bone disease

Hypocalcaemia and hypomagnesaemia are common and often associated with neuromuscular symptoms, dehydration and other electrolyte abnormalities. Osteomalacia is common but difficult to confirm without a bone biopsy. These patients have bone pain and an elevated serum alkaline phosphatase activity.

Treatment

The treatment of short gut syndrome is centred on supporting the patient during the initial stage of decompensation and through the critical stage of adaptation which may last up to 3 months. At the end of this time the stage of equilibrium is reached and the patient with residual small intestine enters the final stage of rehabilitation. In the patients with massive small bowel resection indefinite total parenteral nutrition via a permanent tunnelled silicon feeding line is the only option for survival. In these patients a programme of training with regard to the management of the intravenous feeds and the care of the feeding lines is essential so that they can eventually carry out the parenteral nutrition themselves in their own homes usually at night.

In the immediate postoperative period total parenteral nutrition is essential and the regimen must provide 40 kcal per kilogram of body weight and 300 mg of nitrogen per kilogram of body weight, in addition to electrolytes, vitamins and trace metals, usually in a 3 litre single bag system administered over a period of 12 hours. Accurate fluid balance must be maintained by daily charting of the input versus the output which should include all measured losses. H_2-receptor blockade therapy (cimetadine or ranitidine) should be administered 6 hourly by the intravenous route to suppress gastric secretion. This avoids the necessity for nasogastric intubation and suction which should only be used if ileus is prolonged. Initially nothing other than weak hypotonic electrolyte solutions are tolerated orally and only small amounts (15–30 mL/h) should be given to avoid a dry mouth.

When the patient's condition becomes stable (usually 3–6 weeks after the enterectomy) transition to an oral diet is started in those patients with an adequate length of small intestine. The feeding is started gradually, initially with isotonic carbohydrate and electrolyte solutions. Thereafter elemental diets are used. These are semisynthetic fibre-free liquid diets containing all the basic components (protein, hydrolysates, simple carbohydrates, essential lipids, vitamins). The nitrogen component is available either as free amino acids (elemental) or as peptides (polymeric). There is some evidence that the peptide-based polymeric diets lead to better nitrogen utilization than the elemental diets although this is controversial. The fats which contribute only a small amount of the caloric content of these artificial diets are necessary particularly as a supply of essential fatty acids.

The diet is administered via a nasogastric feeding tube, initially in a dilute form (1:4) and instilled by an infusion pump at a rate of 25 mL/h. The rate and concentration are gradually increased until the patient is receiving 100–120 mL of full strength diet per hour. Care must be taken that gastric dilatation does not occur as pulmonary aspiration with a fatal outcome may occur during nasogastric enteral feeding.

One of the most common early problems is diarrhoea which is often accentuated at the start of enteral feeding. The most effective drugs to control this are loperamide hydrochloride, diphenoxylate hydrochloride and codeine phosphate. In patients with an intact colon, cholestyramine can be administered to limit the diarrhoea-inducing effects of the unabsorbed bile salts. In some patients the diarrhoea is so severe that it may be life threatening and requires control using an infusion of a long-acting analogue of somatostatin.

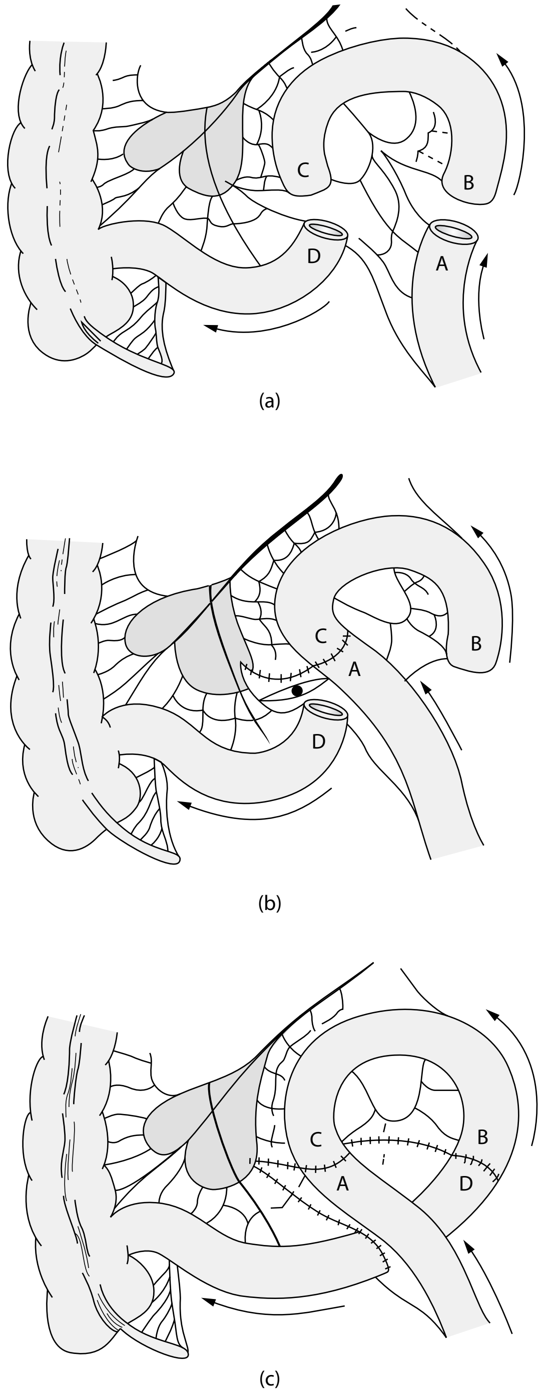
Steatorrhoea is common in these patients and results from excess fat in the diet. Whenever possible, dietary fats should be administered as MCTs which should not require the presence of salts for their absorption into the portal venous blood. Dietary supplements of calcium, magnesium, vitamins C, D and K and iron are all necessary. In patients in whom the ileum has been resected vitamin B_{12} is administered parenterally at 3 monthly intervals on an indefinite basis.

Currently home total parenteral nutrition is the only option for the majority of patients with insufficient residual small bowel. This is costly and attended by significant complications.

Although still in its infancy, combined hepatic and small bowel transplantation offers the best prospects for management of these patients in the long term.

Remedial surgical procedures

In adults, surgery is rarely indicated in patients with short gut syndrome unless intractable diarrhoea proves resistant to all forms of medical treatment including somatostatin. Under these cases reversed (antiperistaltic) segments can be employed. Although they do delay transit they can lead to intestinal obstruction and favour the development of bacterial overgrowth. In order to avoid risks of strangulation the antiperistaltic segment should be constructed without reversal of the mesenteric pedicle (Figure 29.31). In neonates and infants, the procedure of intestinal lengthening has given good results but the procedure is difficult and should be done in expert paediatric surgical centres (Figure 29.32). Truncal vagotomy and pyloroplasty although often used in the past is not indicated as patients who develop persistent gastric acid hypersecretion can be treated using oral omeprazole on a long-term basis.



(a)

(b)

(c)

Figure 29.31 Technique of creating an antiperistaltic segment without reversal of the mesenteric pedicle.

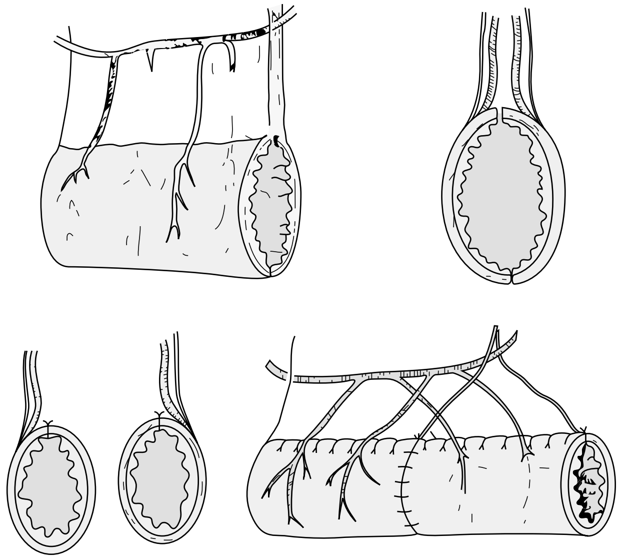


Figure 29.32 Diagrammatic representation of the intestinal lengthening operation of Bianchi. The bowel is split longitudinally into two halves, each retaining its blood supply. A tube is constructed from each half. The two tubes are then joined together.

Coeliac disease (gluten-sensitive enteropathy)

Coeliac disease is the most common and important cause of malabsorption in Western societies, but its prevalence varies considerably. In the UK, the overall prevalence in the population is about 1 in 300. An increased prevalence is also found in individuals with Irish, Punjabi and other South Asian ethnic origin. Coeliac disease is the result of an intestinal allergic reaction to gluten, a dietary protein found mainly in wheat and to a lesser extent in rye and barley. Coeliac disease is an inherited disease of the intestinal immune system and affects both sexes. It is often missed early on in life, being misdiagnosed with other intestinal disorders, such as irritable bowel syndrome and Crohn's disease. Thus the majority of patients are usually diagnosed in adult life. Untreated coeliac disease is associated with long-term health risks, e.g. miscarriages, osteoporosis, anaemia and gastrointestinal malignancy.

Pathology

Coeliac disease is characterized by a mucosa with shortened villi and lengthened crypts in the upper small intestine. The severity of the disease increases with the extent of involvement of small intestine but it is relatively unusual for the whole of the ileum to be involved. The histological hallmark of coeliac disease is subtotal villous atrophy but the mucosa remains of normal thickness due to the crypt hyperplasia.

Aetiology

Coeliac disease is regarded as a multigenetic disorder, associated with HLA types HLA-DQ2 (90%) or HLA-DQ8, plus other genetic or environmental factors. There is a familial tendency with 10–15% of first-degree relatives being affected. Antenatal HLA testing is used to provide information to parents of a coeliac disease child desiring to know the risk of the disease in a further child. The disease is also commoner in patients with type 1 diabetes mellitus (2–8%), Down syndrome, sarcoidosis, infertility (both sexes), patients with IgA antibody deficiency and certain autoimmune disorders (thyroid disease, rheumatoid arthritis, chronic active hepatitis, Addison disease and Sjörgen syndrome).

Clinical features

The disease may present at any time of life from the first few months to old age. Coeliac disease presentation in childhood is uncommon in the UK with only 9% being diagnosed in this age group. Babies and young children present after weaning (9 months to 3 years) with malabsorption and diarrhoea (frequent pale stools) but sometimes paradoxically with constipation, refusal to feed, weight loss/failure to thrive, vomiting, anorexia and irritability. On examination, the abdomen is protruberant with eversion of the umbilicus.

Older children and adults may present with anaemia (folate or iron deficiency), abdominal discomfort, arthralgia, fatigue and malaise, diarrhoea, steatorrhoea and malabsorption. On examination, mouth ulcers and angular stomatitis are common and are associated with deficiencies of folate (85%), vitamin D (up to 30%), vitamin K (10%) and vitamin B_{12}. Some adult coeliac disease patients present with infertility (both sexes) and females with repeated miscarriages.

The classical manifestations of coeliac disease include malabsorption, chronic diarrhoea, anaemia, rickets, failure to grow, abdominal bloating, offensive bulky stools, dermatitis herpetiformis (blistering itchy rash on the back, elbow and scalp) and mouth ulcers. In adults, coeliac disease often presents with non-specific symptoms, e.g. fatigue, mouth ulcers, skin rashes, vague abdominal pains, intermittent diarrhoea and chronic anaemia. The vast majority of patients who develop dermatitis herpetiformis will have coeliac disease.

Diagnosis

Antibody screening tests

These measure antibodies to gluten or gliadin and endomysial antibodies (EMAs) to the damaged bowel wall and are used as the initial screening test. The antigliadin antibodies, which disappear on a gluten-free diet, are no longer recommended as they are less specific and can be positive in other gastrointestinal conditions, such as Crohn's disease.

Assays for IgA antitissue transglutaminase antibody (tTGA), which is the autoantigen of EMA, are nowadays the preferred investigation. EMAs are used if the tTGA test is not available or equivocal. Both tests are highly specific and sensitive for untreated coeliac disease, provided the patient is still on a gluten diet. Even so in patients with positive antibody tests the disease needs to be confirmed by an intestinal biopsy performed with the patient on a normal (wheat) diet. False-negative results occur in patients with selective IgA deficiency (0.4% of the general population). The current recommendation is for the use of IgG tTGA and/or IgG EMA serological tests in patients with confirmed IgA deficiency. Antibodies frequently become undetectable after 6–12 months on a gluten-free diet and can be used to monitor the disease.

Serological testing is recommended in patients with irritable bowel syndrome and children with type 1 diabetes. The National Institute for Health and Clinical Excellence also recommends serological testing in patients with Addison disease, amenorrhoea, aphthous stomatitis, autoimmune liver disease, autoimmune myocarditis, chronic thrombocytopenic purpura, dental enamel defects, depression, Down syndrome, epilepsy, lymphoma, metabolic bone disease (rickets or osteomalacia), persistent or unexplained constipation, polyneuropathy, recurrent miscarriage, sarcoidosis, Sjögren syndrome, Turner syndrome, unexplained alopecia and unexplained subfertility.

Other tests

A small bowel follow-through shows an abnormal appearance in 90% of patients although the appearances are non-specific. The normal fine feathery appearance of the mucosa is no longer seen but is replaced by a coarse mucosal pattern with broad bars. As the disease becomes more severe the small bowel may appear as a featureless tube. Changes are mainly seen in the jejunum but distal spread is seen in more severe disease.

The most important diagnostic test is jejunal biopsy which can be achieved either at upper gastrointestinal endoscopy or using a suction (Crosby) capsule. At endoscopy it is common to see loss of the folds of Kerckring in the descending duodenum and this should alert the endoscopist to the possibility of coeliac disease. In any case, the diagnosis of coeliac disease is confirmed when biopsy shows villous atrophy while the patient is on a gluten diet, followed by full clinical remission with a gluten-free diet when the tTGAs or EMAs gradually disappear. Further biopsy following gluten challenge may be needed in borderline cases.

Other investigations which should be done include full blood count and blood film for Howell–Jolly bodies (splenic atrophy), liver function tests and small bowel contrast studies to exclude other causes of malabsorption/diarrhoea or lymphoma.

Complications

To a certain extent, although the gut is the main organ of gluten sensitivity, the condition can be systemic. The non-bowel-related complications include dermatitis herpetiformis and neurological dysfunction leading particularly to cerebellar ataxia often with a peripheral neuropathy. The main complications of coeliac disease related to the intestinal disease are nutritional deficiencies leading to iron-deficiency anaemia, polyvitamin deficiencies and osteoporosis with fractures (50%). The patients are also at an increased risk of mouth, throat, oesophageal and small bowel cancer. Coeliac disease is also associated with an increased risk of developing some types of lymphoma, the commonest being enteropathy-type T-cell lymphoma but these patients have an increased risk of other types of NHLs. Overall the incidence of malignancy in adult coeliac disease is in the region of 10%.

Treatment

The only effective treatment of coeliac disease is strict lifelong complete avoidance of gluten found in cereals such as wheat, rye and barley (many tolerate oats). It is completely safe for coeliac disease patients to eat maize, corn or rice. With the increased potential for developing osteoporosis, coeliac disease patients should be closely monitored and all menopausal women and men over 55 years of age who suffer with coeliac disease should have bone density (DEXA) scanning.

The gluten-free diet rapidly induces clinical improvement, which is mirrored by restoration of the villous architecture of the mucosa. The diet consists of no wheat, barley, rye, or any food containing them. Although moderate quantities of oats are allowed, the British Society of Gastroenterology guidelines suggest oats be excluded at least for the first year while patients get accustomed to a gluten-free diet before oats are cautiously introduced. A gluten-free diet should be life-long, as relaxation of diet precipitates a relapse of symptoms and increases the incidence of complications. The diet should be supplemented by folic acid, iron, calcium and vitamin D as necessary. Serial tTGAs or EMAs are used to monitor response to diet. tTGA is the autoantigen of EMA.

Vascular abnormalities of the gut

These constitute a heterogeneous group of conditions which share a related pathology consisting of abnormal lesions of the gut alone or with similar lesions in other organs including the skin. Not all lesions are symptomatic, but when they are any of the gut lesions present most usually with obscure recurrent gastrointestinal bleeding and less commonly with life-threatening haemorrhage. In the vast majority of cases, treatment is with endoscopic therapy, less commonly with interventional angiography. Surgery has a limited role in the treatment nowadays, being reserved for major bleeding which is not controlled by other therapies. The lesions which are malformations rather than neoplasms include angiodysplasia, phlebectasia, telangiectasia and haemangiomas.

Angiodysplasia

The terms angiodysplasia, arteriovenous malformation, angioectasia, and vascular ectasia are often used synonymously. The confusion is added to by the fact that while some authors use angioectasia as a generic term, others reserve it for colonic lesions.

Angiodysplasias are lesions which occupy the mucosa and submucosa of the gastrointestinal tract and consist of a cluster of arteriolar, venular and capillary vessels. Although this condition can occur in the small bowel it is most commonly seen in the right side of the colon. Angiodysplasia may cause gastrointestinal bleeding usually of the recurrent occult type, but rarely the bleeding may be overt and severe, causing life-threatening hypovolaemia. The aetiology is uncertain but theories include chronic mucosal ischaemia secondary to arteriovenous shunting, decreased perfusion pressure and lowered oxygen tension in the terminal branches of the SMA and chronic intermittent obstruction of the submucosal veins.

Pathology and aetiology

Gastrointestinal angiodysplasia is usually found in the elderly (≥60 years) with an estimated incidence of 0.8%, but may rarely affect younger patients. However, the prevalence of angiodysplasia is increased in patients with advanced renal disease, aortic valve stenosis and in von Willebrand's disease. Approximately half of affected patients have more than one lesion. The majority of angiodysplasias occur in the colorectal region, mainly in the right colon (70–80%), the commonest site being the caecum. However, lesions in the small bowel and gastroduodenal region may occur in their own right, or as synchronous lesions in association with colonic angiodysplasias.

Angiodysplasias are composed of ectatic, thin-walled vessels lined by endothelium alone or with small amounts of smooth muscle intimately associated with dilated, tortuous submucosal veins. Large lesions are associated with enlarged feeding arterioles and arteriovenous fistulas. The pathogenesis of angiodysplasia remains unclear. The most favoured theory is intermittent, low-grade obstruction of submucosal veins at the level of the muscularis propria. In time this venous obstruction results in dilatation and tortuosity of the related submucosal venules and capillaries. This hypothesis is consistent with the high incidence of angiodysplasia in the right colon where intraluminal and wall tension is highest. Another hypothesis postulates increased expression of angiogenic factors as the cause. A possible mechanism by which aortic stenosis may lead to the development of angiodysplasia is through the development of an acquired form of von Willebrand disease due to mechanical disruption of von Willebrand multimers from the turbulent blood flow passing through the narrowed aortic valve in association with the thrombocytopenia which characterizes this disease.

Angiodysplasia of the small intestine presents with recurrent obscure bleeding in patients with negative upper gastrointestinal endoscopy and colonoscopy (5%). These cases are nowadays diagnosed with small bowel endoscopy or DBe which can also be used for treatment. Angiodysplasia of the stomach or duodenum is the least common and is responsible for 5% of gastrointestinal bleeding episodes, which may be occult, presenting with iron-deficiency anaemia, or overt, causing haematemesis.

Clinical manifestations and diagnosis

The usual presentation of angiodysplasia is with recurrent/chronic obscure gastrointestinal bleeding, which is usually minor but may rarely be severe in patients with large arteriolized lesions and accompanied by life-threatening hypovolaemia. Angiodysplasia may also be discovered incidentally during an endoscopy for unrelated reasons.

In patients with end-stage renal disease, angiodysplasia is the cause of upper and lower gastrointestinal bleeding recurrent episodes in 30%. In these patients angiodysplastic lesions are often multiple and may occur anywhere in the gut. The increased risk of bleeding in uraemic patients is thought to be platelet dysfunction induced by the end-stage renal disease. The higher incidence of bleeding from angiodysplasia in patients with von Willebrand disease is related to the underlying coagulopathy. In patients with aortic stenosis the recurrent bleeding episodes are reduced markedly in those treated with aortic valve replacement.

Angiodysplasia is usually diagnosed by endoscopy (upper gastrointestinal, colonoscopy, CE, single/double balloon enteroscopy) performed in patients with episodes of gastrointestinal bleeding. Initially all patients have upper gastrointestinal endoscopy and colonoscopy, with other endoscopy tests being performed if the cause of the gastrointestinal bleeding remains obscure. Endoscopically, angiodysplasias have a characteristic appearance of small (5–10 mm), flat, cherry red lesions with a mesh of ectatic blood capillaries radiating from a central vessel. Opioid antagonist (naloxone) administration is reported to increase the detection rate of colonic angiodysplasias. CT or magnetic resonance angiography are also used for the diagnosis of angiodysplasia, although their sensitivity and specificity are lower than that of colonoscopy. In angiodysplasia the typical angiographic finding consists of a tortuous feeding artery and dilated veins with an intervening cluster of fine vessels. When there is active bleeding angiography may demonstrate extravasation of contrast into the small bowel lumen and scanning using labelled autologous red cells may be of value.

Treatment

The current opinion is that angiodysplasias discovered incidentally during colonoscopy do not require active treatment as the risk of subsequent bleeding is low. In contrast, all patients with occult bleeding require treatment as the risk of rebleeding is high especially in patients with multiple lesions and in patients with a bleeding tendency from any cause. All patients with overt bleeding require immediate treatment.

A variety of endoscopic treatments can be used. Argon plasma coagulation is the most common modality as it is safe, effective and cheap. Bipolar electrocoagulation and heater probe coagulation are also effective in the treatment of angiodysplasia, although perforation with heater probe coagulation is a recognized risk. Monopolar electrocoagulation is no longer used because of this risk. Other techniques for arrest of bleeding angiodysplasias include mechanical haemostatic methods with endoscopic clips and band ligation. These are favoured in patients on anticoagulants and/or antiplatelet agents, and in patients with coagulation defects. Angiography to localize the site of active bleeding followed by its embolization-is reserved for patients with life threatening bleeding who are not surgical candidates. Surgery has a limited place and is only considered in patients with large transfusion requirements or life-threatening haemorrhage from a clearly identified site. For lesions of the small bowel causing significant haemorrhage, resection of the affected segment of small bowel is carried out. Ideally, if mesenteric angiography has been carried out prior to surgery, the radiologist should leave a superselective catheter in a mesenteric vessel as close to the lesion as possible. At surgery, the surgeon can then inject some dye (methylene blue) into the catheter and this will highlight the segment of small bowel to be resected.

Phlebectasia

Phlebectasias are rare benign vascular anomalies of the gastrointestinal tract, with only a few reported cases to date. These lesions are usually asymptomatic and are usually discovered at postmortem or incidentally during surgery, but, occasionally, they present with mild to massive gastrointestinal haemorrhage. Essentially they consist of a meshwork of dilated veins having a normal endothelial lining and situated in the submucosal layer of the intestine with a thin overlying mucosa. Endoscopically, they appear as multiple, dark bluish red nodules, less than 1.0 cm in diameter, which blanch on pressure. The exact cause of phlebectasia is not known and suggested theories include a congenital abnormality of the endothelial lining of submucosal veins and age-related loss of connective tissue support surrounding the submucosal veins.

Phlebectasia, which may be multiple, can occur in any segment of the gastrointestinal tract, including the small intestine but the jejunum is one of the commonly reported sites. When discovered, phlebectasia must be distinguished from varices as the endoscopic appearances are identical, distinction being based on the absence of cirrhosis and other manifestations of portal hypertension. Histologically, they consist of dilated veins lined by a single layer of endothelial cells and irregular bundles of smooth muscle in the tunica media without internal or external elastic lamina. Detection can be difficult in symptomatic cases but the same investigations described in the section on angiodysplasia are used to establish the diagnosis. Multiple phlebectasia of the jejunum, associated with similar lesions of the oral mucosa, tongue ('caviar spots') and scrotum (Fordyce lesion) have been reported.

Hereditary haemorrhagic telangiectasia (Osler–Weber–Rendu disease)

Hereditary haemorrhagic telangiectasia (HHT) is a rare inherited disorder which leads to multiple telangiectasias of the lip, mouth, nasopharynx and gut and other internal organs. The lesions consist of arteriolar dilations and arteriovenous malformations due to a congenital weakness of the arterial muscle and absence of elastin in the medial coat. The prevalence of HHT is estimated at one in 5000–8000 individuals. The coagulation is normal in affected patients. HHT, also known as Osler–Rendu–Weber disease, is characterized by a tendency for bleeding, especially recurrent epistaxis, cutaneous telangiectasia and bleeding in internal organs including the small intestine.

Pathology

HHT is an autosomal dominant vascular dysplasia leading to telangiectasias and arteriovenous malformations of skin, mucous membranes and internal organs. Two main types of HHT are recognized: (1) the most frequent (HHT1) is due to mutations of the *END* gene coding for endoglin on the long arm of chromosome 9 at 9q34.1 (MIM.187300); and (2) HHT2 due to mutations at 12q11-q14 (MIM.600376) coding for the *ACVRL1* gene. There is also a variant in which HHT occurs in association with juvenile polyposis.

Clinical features

The diagnostic criteria for HHT include (1) spontaneous recurrent nosebleeds, (2) multiple telangiectasias on skin and mucous membranes, (3) involvement of internal organs and (4) an affected parent, sibling or child. The first sign of HHT usually does not occur until puberty or adulthood with the average age of the first nosebleed occurring at 12 years. Recurrent nosebleeds are seen in 50–80% of patients with HHT. Bleeding episodes vary from daily to weekly and monthly nosebleeds depending on severity of the disease – roughly one-third mild, one-third moderate and one-third severe. Telangiectasia of the skin and mucous membranes may occur anywhere but predominate on the upper half of the body including the face, inside the mouth and nostrils, lips, ears, conjunctiva of the eyes, forearms, hands and fingers. Telangiectasia and arteriovenous malformations in other parts of the body affect 95% of patients but these usually appear in adult life (between 20 and 40 years). Lesions can be found anywhere in the Gastrointestinal tract (oesophagus, stomach and small and large bowel). Gastrointestinal bleeding occurs in about 25% of patients with HHT and the risk increases in patients older than 50 years. Other organs which may be involved with internal bleeding include the lungs (30%), the brain and spinal cord.

Treatment

Mild cases usually require no treatment. HHT requires treatment if it causes significant problems, such as severe and/or frequent nosebleeds, or if there is a high risk of causing other problems. Nosebleeds are treated with laser coagulation therapy or surgically with nasal septal dermoplasty or complete nasal closure (Young's procedure). Gastrointestinal bleeding causing anaemia is treated with iron replacement therapy. If this is ineffective then blood transfusion and endoscopic coagulation are needed.

Gastrointestinal haemangiomas

Gastrointestinal haemangiomas are uncommon benign vascular malformations that may occur anywhere in the gastrointestinal tract as single or multiple lesions. They are commonest in the small intestine accounting for 10% of small bowel lesions, followed by the colon. Multiple lesions are often associated with haemangiomas in other organs (liver and skin) and may also be a feature of other disorders, e.g. *Osler–Weber–Rendu disease*, *Maffucci syndrome*, *Klippel–Trénaunay syndrome*, or the congenital *blue rubber bleb naevus syndrome*. They present most commonly with gastrointestinal bleeding, which may be slow and occult or overt with life-threatening hypovolaemia.

Pathology

Macroscopically gastrointestinal haemangiomas usually form pedunculated polypoid masses and less commonly diffuse submucosal infiltrating lesions. The common sites in descending order of reported cases are the small intestine (40%), colon (30%) and gastro-oesophageal (10%). Very rarely, they are diffuse throughout the gastrointestinal tract and may involve other organs (gastrointestinal haemangiomatosis). Haemangiomas appear endoscopically as red to bluish lesions with a soft compressible consistency, unless they contain areas of thrombosis/calcification (phleboliths). Histologically, they are classified in accordance with the nature of their major vascular components into three types: capillary, cavernous or mixed. Capillary haemangiomas consist of a mass of small capillaries and thin-walled blood-filled spaces, whereas the cavernous types contain large sinuses lined by single or multiple layers of endothelial cells. Cavernous haemangiomas may infiltrate large segments of the small intestine and adjacent mesentery. Both types may exhibit areas of thrombosis with calcification and hyalinization. When haemangiomas originate from the mesentery, large segments of the small bowel may be involved. Focal calcification, thrombi and hyalinization represent degenerative changes.

Clinical features

Irrespective of site of origin, gastrointestinal bleeding is the most common clinical presentation. The bleeding may be slow and insidious or massive and life threatening. Cases have been reported when extensive cavernous haemangiomas have been complicated with diffuse intravascular coagulation with widespread massive bleeding due to the coagulopathy. Patients with small bowel lesions may also present with abdominal pain, mechanical bowel obstruction due to the formation of intramural haematoma, intussusception or perforation. The differential diagnosis of small bowel haemangiomas includes other benign and malignant small intestine tumours as well as metastatic disease. Other conditions, e.g. lymphoma, metastatic disease, primary peritoneal malignancies and inflammatory pseudotumour, should be considered in the differential diagnosis for lesions that infiltrate the mesentery and small intestine.

In patients with extensive involvement, small bowel contrast radiology may show displacement of the bowel, irregularities of the mucosal folds or a polypoid intraluminal mass. CT can identify a large mass contiguous with the small intestine in these patients, but small lesions may be radiographically occult and difficult to detect before surgery. Abdominal plain radiographs may show phleboliths. Occasionally, when haemangiomas originate from the mesentery, large segments of the small intestine may be identified as a large mass contiguous with the small intestine on CT. In contrast, small lesions may be missed and these are notoriously difficult to detect by radiological tests before surgery. However, except in the emergency situation, these investigations have been largely replaced by CE and balloon endoscopy (single and double). CE is usually done first as it is far less invasive. However balloon endoscopy gives a more precise assessment of the extent of the disease and in selected cases (e.g. small solitary polypoid lesions) may enable endoscopic treatment.

In the colon the most common site is the rectosigmoid followed by the right colon and caecum. Most patients with rectosigmoid haemangiomas are young men. Irrespective of colonic segment involved, all patients present with rectal bleeding. The presence of clusters of phleboliths with atypical distribution within the pelvis is common in colorectal haemangiomas and is a useful sign in young patients. Haemangiomas of the rectosigmoid may widen the retrorectal space on CT caused by the transmural thickening of the involved segment. MRI of rectosigmoid haemangiomas shows focal thickening of the colonic wall with high signal intensity on T_2 weighted images in both the lesion and perirectal fat.

Oesophageal haemangiomas constitute 3.0% of all benign oesophageal tumours. They may occur at any level in the oesophagus and often present with dysphagia, but may also produce haematemesis and melaena. Diagnosis is established by barium swallow which outlines a well-defined lobulated or a pedunculated intraluminal mass and upper gastrointestinal endoscopy. When discovered at endoscopy, biopsy must never be attempted because of the real risk of catastrophic haemorrhage. The differential diagnosis.

Treatment

Once diagnosed all gastrointestinal haemangiomas require treatment in view of their propensity to bleed in addition to other complications. It is best to consider treatment in the elective and emergency situations.

Elective treatment

With the exception of small polypoid lesions which can be treated endoscopically by balloon endoscopy (laser fulguration, heater probes, etc.), provided local expertise is available, the vast majority of small bowel lesions require segmental resection of the affected small intestine. This is usually carried out by open surgery but, provided the lesion is accurately localized, there is no reason why this cannot be done by the laparoscopic approach. Indeed there have been several reported series of the value and benefit of laparoscopy both in the diagnosis and accurate localization of small bowel haemangiomas and in the surgical treatment by laparoscopic segmental resection. In most of the reported series, the enteroenteric anastomosis has been done extracorporeally though the small incision used to extract the specimen, but internal restoration of continuity by stapling or intracorporeal suturing is a valid alternative.

Emergency treatment

Patients presenting with acute massive gastrointestinal haemorrhage require replacement of the hypovolaemia and stabilization with blood/colloid transfusions. This is followed, soon after, because of the real risk of recurrence of bleeding, with emergency radiological intervention requiring superelective cannulation of the feeding vessels to locate and embolize the lesion. Often a blue dye is also injected together with the contrast medium. The reason for this is that should the embolization fail and an emergency laparotomy is needed, the surgeon can more easily locate the involved small bowel segment. Less commonly, emergency laparotomy is need for haemangiomas causing perforation or small bowel obstruction from intussusception or the formation of intramural haematoma. Again resection with end-to-end anastomosis is carried out.

Gastrointestinal haemangiomatosis

This is a complex vascular malformation that occurs primarily in infancy and childhood. Most patients present with gastrointestinal bleeding. In others, the presentation is with intussusception, small bowel obstruction, perforation or protein-losing enteropathy/malabsorption. Gastrointestinal haemangiomatosis may be associated with blue rubber bleb naevus syndrome, Klippel–Trénaunay–Weber syndrome, Maffucci syndrome, diffuse neonatal haemangiomatosis and Proteus syndrome. Pathologically, haemangiomatosis is characterized by diffuse infiltration of the intestinal wall, the mesentery, and, occasionally, the retroperitoneum. Solid organs in the abdomen may also be involved. The radiographic findings include the presence of phleboliths (plain abdominal films) and mural thickening on CT.

The vermiform appendix and Meckel's diverticulum

Vermiform appendix

Anatomy

The appendix appears during the fifth month of gestation. Initially several lymphoid follicles are scattered throughout the developing organ. These follicles increase with growth up to 8–20 years of age. The vermiform appendix is a narrow (around 5 mm) diverticulum of highly variable length (5–15 cm). The base of the appendix is attached to the posteromedial surface of the caecum approximately 3 cm inferolateral to the junction of the ileum and caecum. Its blood supply, which originates from the ileocolic artery, reaches it via a small extension of the mesentery of the small bowel, which passes posterior to the terminal ileum. The appendix is highly variable in position. Often (65%) it lies posterior to the caecum (retrocaecal) but it may also extend into the iliac fossa and lesser pelvis (30%) and lie close to the ovary, fallopian tube and ureter. As with the small bowel, the appendix is covered with peritoneum and has an external layer of longitudinal muscle and an internal layer of circular muscle. At the base of the appendix the longitudinal muscle is continuous with the three taeniae of the caecum and the colon. In the submucosal layer there are lymph follicles which are separated by crypts of the columnar epithelial lining that contains many goblet cells. When the submucosal layer becomes swollen it can readily block the lumen of the appendix.

Aetiology and epidemiology of appendicitis

Appendicitis is caused by obstruction of its lumen, most commonly by lymphoid hyperplasia (more common during childhood and in young adults) secondary to various inflammatory and infectious conditions, e.g. gastroenteritis, respiratory infections, measles, infectious mononucleosis and inflammatory bowel disease. The other common cause of appendiceal obstruction leading to appendicitis is faecalith impaction (more common in elderly patients). Rare causes include parasites, foreign bodies and neoplasms. Obstruction is less commonly due to bacterial infections.

The high incidence of acute appendicitis in the West and its relative rarity in Africa suggest a protective effect of a high-fibre diet. The high-fibre diet decreases bowel transit time and is thought to minimize the formation of faecaliths. There is also evidence that appendicitis is commoner in urban society than in rural districts, and this may be attributed to a high incidence of enteric infections related to crowded living conditions. Domestic and food hygiene have improved dramatically over the past 40 years and this has coincided with a significant fall in the incidence of appendicitis. There is also epidemiological evidence indicating that the consumption of green vegetables and tomatoes may be protective against appendicitis, whereas potato consumption appears to be related to the disease. In elderly patients there is some evidence that chronic intake of non-steroidal anti-inflammatory drugs may increase the risk.

Acute appendicitis

Pathology

The obstruction causes an elevated intraluminal pressure from increased secretion of fluids and mucus by the mucosa with stasis of the contents and proliferation of intestinal bacteria which recruit polymorphonuclear leucocytes forming pus, thereby inducing a further rise in the intraluminal pressure. This ultimately leads to venous outflow obstruction, ischaemia, loss of epithelial integrity and bacterial invasion of the appendiceal wall. As the pathology progresses thrombosis of the appendicular artery and veins occurs with gangrene and perforation of the appendix forming a periappendicular abscess or local/generalized peritonitis.

The presence of gangrene or perforation seems to be associated with the presence of faecaliths, which are radio-opaque intraluminal laminated appendiceal calculi. Approximately 50% of cases of gangrenous or perforated appendicitis are associated with a faecalith in contrast with uncomplicated appendicitis in which a faecalith is rarely present. The rate of perforation varies from 16% to 40%, with a higher frequency occurring in younger age groups (40–57%) and in patients older than 50 years (55–70%), in whom misdiagnosis and delayed diagnosis are common.

Clinical features

Appendicitis is one of the more common surgical emergencies and is the commonest cause of abdominal pain requiring hospital admission. In this context, acute abdominal pain is defined as previously undiagnosed pain of <72 hours' duration. The current annual incidence of acute appendicitis is one in 1000 individuals. There is a slight male preponderance (3:2) in teenagers and young adults. In adults, the incidence of appendicitis is approximately 1.4 times higher in males. The incidence of appendicitis peaks in the late teen years, and gradually declines thereafter. In the paediatric age group the mean age at presentation is 6–10 years. This is thought to result from the higher prevalence of lymphoid hyperplasia.

History

Owing to the embryological origin of the appendix as a midline structure the majority of patients with acute appendicitis first notice a pain which starts in the region of the umbilicus. This pain may be a dull ache or it may be colicky pain presumably owing to obstruction of the appendiceal lumen. After a variable period of time the pain shifts to the right lower quadrant of the abdomen owing to the inflamed appendix irritating the parietal peritoneum. It must be stressed, however, that approximately 30% of patients do not experience this shift of pain and their symptoms start with discomfort in the right lower quadrant. Patients will usually report that movement, particularly coughing or laughing, leads to sharp exacerbations of the pain. Nausea and vomiting are common and anorexia is almost inevitable. It is important to appreciate that about 20% of patients will also have diarrhoea, particularly when the appendix lies in the pelvis, and this may lead to a mistaken diagnosis of gastroenteritis. Symptoms may be altered by other location of the inflamed appendix: flank and back pain in retrocaecal appendicitis, urinary symptoms with pelvic location.

In taking the history it is important to ask questions which might alert one to possible alternative diagnoses. Particularly in children it is important to ask about a sore throat or flu-like symptoms as these often accompany mesenteric adenitis. The patient should be asked about dysuria, frequency and cloudy or strong smelling urine, as urinary tract infection can often lead to lower abdominal pain. In women, a menstrual history is essential. A missed period may point to an ectopic pregnancy. Pain at mid-cycle may indicate that the pain is due to ovulation (Mittelschmerz) and vaginal discharge may indicate pelvic inflammatory disease.

Examination

The first part of the examination should consist of general inspection of the patient's wellbeing with measurement of their pulse and temperature. Patients with appendicitis may well have a normal pulse rate and temperature, but a sustained tachycardia is highly significant in the presence of abdominal pain and tenderness and should always be taken seriously. Although the temperature may be normal, most patients with appendicitis have low-grade pyrexia. A very high temperature (above 39°C) indicates probable abscess formation or some other diagnosis such as a viral illness.

The next stage is to observe the patient's abdomen and to look for movement with respiration. It is then useful to ask the patient to cough while watching their facial expression. If coughing produces obvious pain the patient should be asked to indicate the site of maximum pain. If this lies in the right iliac fossa this will indicate localized peritonitis in this area and is highly suggestive of appendicitis. The tongue, mouth and throat should be examined as a furred tongue and fetor oris will be present in about 50% of patients, particularly those with a delayed diagnosis. The tonsils should also be inspected particularly in children, as tonsillitis may be associated with mesenteric adenitis.

Attention should then be turned to palpation of the abdomen. The first point to establish is the site of maximal tenderness. In the majority of patients this is at or close to 'McBurney's point' which is situated at the junction between the upper two-thirds and lower one-third of a straight line joining the umbilicus and the anterior superior iliac spine. It must be stressed, however, that in patients with inflammation in a retrocaecal appendix the pain may be considerably higher and more lateral than this, and in pelvic appendicitis the pain may be lower and almost midline. Indeed, with a low pelvic appendix tenderness may be only detectable by rectal examination. The abdomen should then be assessed for guarding or the involuntary contraction of the abdominal wall muscles over the area of inflamed peritoneum. Right lower quadrant guarding is found in about 90% of patients with acute appendicitis and if the appendix has perforated leading to generalized peritonitis the area of guarding may extend beyond the right lower quadrant. Rebound tenderness is another useful sign. This may be elicited by pressing gently on the right lower quadrant of the abdomen and then suddenly releasing the hand and watching the patient's face for signs of discomfort. Another approach is to use percussion which, in the presence of peritoneal irritation, will elicit the same response and is kinder.

When the amount of tenderness permits, careful palpation of the right iliac fossa for a mass should be carried out. This may indicate the presence of an appendiceal abscess or an appendix mass (phlegmon) created by omentum wrapped around the inflamed appendix. Of course, it may also indicate some other pathology such as a caecal carcinoma which is mimicking appendicitis in its presentation.

After careful examination of the abdomen, attention should be turned in the male to the testes as both acute torsion and orchitis may present initially with right lower quadrant pain. The hernial orifices should also be carefully examined for strangulated inguinal or femoral hernias and it should never be forgotten that acute appendicitis can occasionally occur in an appendix lying within a hernial sac.

Rectal examination can be extremely helpful in the clinical assessment of suspected appendicitis but in a patient with clear-cut symptoms and signs it is unnecessary. However, particularly in a patient with diarrhoea and abdominal pain who does not have convincing abdominal signs, a rectal examination should always be carried out. The examining finger should be pressed on to the pelvic peritoneum first to the left and then to the right. When there is a pelvic appendicitis there will be more discomfort felt on the right. In female patients, the surgeon should then take opportunity of moving the cervix through the rectal wall as pain associated with this manoeuvre is highly indicative of pelvic inflammatory disease.

Diagnosis

The diagnosis of acute appendicitis is made largely on clinical grounds. The problem is that the diagnosis is correct in only 50% and up to 20% of appendicectomy specimens are normal especially in females. A systemic review of published series identified the following useful predictors of appendicitis in patients with abdominal pain: raised inflammatory markers, clinical signs of peritoneal irritation, migration of abdominal pain. However, certain investigations which may be of value in problematic cases include:

- elevated white cell count and C-reactive protein
- a plain abdominal radiograph
- ultrasound and CT scanning
- pregnancy test in adult females
- laparoscopy
- urinalysis.

The majority of patients with acute appendicitis will have a polymorphonuclear leucocytosis and a white cell count of more than 14×10^9/L is suggestive of appendicitis. However, it must be remembered that various other causes of abdominal pain result in an increased white cell count and around 25% of patients with appendicitis have a normal preoperative white cell count. In patients in whom the diagnosis of acute appendicitis is doubtful and in whom a period of observation is felt to be appropriate, it is worthwhile repeating the white cell count. A falling white cell count will strongly support the diagnosis of non-specific abdominal pain.

In a patient with clear-cut symptoms and signs of acute appendicitis a plain abdominal radiograph is of little or no diagnostic value. However, it is clearly indicated if there is some clinical suspicion of intestinal obstruction or ureteric colic. An erect chest radiograph should always be requested when there is a suspicion of a perforated viscus. A urinalysis should always be carried out to exclude the possibility of urinary tract infection, bearing in mind that about 20% of patients with appendicitis will have sterile proteinuria and pyuria.

Ultrasound and CT scanning

Ultrasound examination of the pelvis is particularly useful in female patients when a differential diagnosis of gynaecological pathology such as a twisted ovarian cyst is being entertained. It may also be useful in distinguishing between an appendix mass and an abscess. In addition to this there have been some reports indicating that ultrasound can be used to make the diagnosis of appendicitis with a high level of sensitivity and specificity. This has not become routine however owing to the fact that a high level of expertise is required. Such expertise is not routinely available for emergency situations in most institutions. CT scanning may occasionally be useful in establishing the diagnosis where a right lower quadrant mass is present. However, it does not have a place in the routine diagnosis of acute appendicitis.

Laparoscopy

Laparoscopy has an established role in the diagnosis of acute appendicitis when the clinical diagnosis is uncertain. It is particularly of value in women of child-bearing age where there are a number of different causes of right lower quadrant pain and tenderness. Several studies now testify to the reduction in the negative appendicectomy rate in such patients. Use of laparoscopy in males and in children is less well established, but whenever there is a diagnostic problem and the surgeon is unhappy about prolonged observation then laparoscopy is appropriate in any patient who has not had previous extensive abdominal surgery. During the procedure a probe should always be used in order to manipulate the caecum so that the appendix can be fully visualized.

Laparoscopy is greatly superior to a right lower quadrant gridiron incision for visualizing the abdominal and pelvic organ. For this reason it is particularly useful in planning incisions both for acute appendicitis and for other unsuspected intra-abdominal pathology. Laparoscopy can also be used to carry out appendicectomy (see section on treatment).

Complications

Complications of acute appendicitis are secondary to gangrenous or perforated appendicitis and can therefore be avoided by prompt recognition of the condition and appropriate treatment. These are:

- appendix abscess
- appendix mass
- wound infection
- generalized peritonitis, particularly in the very young or elderly
- intraperitoneal abscess formation, either subphrenic or multiple small intraloop abscesses
- faecal fistula usually following draining of an abscess
- recurrent intestinal obstruction due to the formation of adhesions
- portal pyaemia
- sterility in women of child-bearing age; there is some debate as to whether perforation of the appendix increases the rate of infertility but recent studies indicated that it is probably not a major except for patients with perforation and peritonitis leading to obstruction of fallopian tubes from adhesions
- overwhelming sepsis and death.

Treatment

In cases of doubt regarding the diagnosis, a period of 'active observation' is advisable. Reports suggest that this approach reduces the negative appendicectomy rate without increasing the risk of perforation. Intravenous fluids and analgesia should be given. Opiate analgesia should be administered as it does not mask the signs of peritonism. Antibiotics should be withheld until the decision to operate has been made.

The mainstay of treatment in acute appendicitis is early appendicectomy. This is normally done through a small skin crease incision in the right iliac fossa, which involves splitting the underlying abdominal muscles rather than cutting them. After removal of the appendix, the appendiceal stump is traditionally buried into the caecum but there have been trials to show that this manoeuvre is unnecessary. If the appendicitis is diagnosed at the time of a laparoscopy it is possible to carry out a laparoscopic appendicectomy if the surgeon has the necessary expertise. There have now been a number of randomized trials comparing conventional appendicectomy with laparoscopic appendicectomy. These have shown similar outcome except for reduced postoperative pain and a lower postoperative wound infection rate with the laparoscopic approach.

In all patients undergoing appendicectomy, prophylactic antibiotics should be used and a combination of metronidazole and cefuroxime is widely favoured. However there is evidence that metronidazole alone, administered as a suppository, is appropriate. In the patient who has a perforated appendix, appendicectomy should be followed by peritoneal lavage with saline containing an antibiotic (cefuroxime or tetracycline). When perforation has occurred it is common practice to continue intravenous antibiotics for 3–5 days postoperatively. Clinical trials have shown that drains do not confer any benefit and may indeed increase the incidence of wound infection.

Appendicitis can be treated conservatively by means of bed rest and intravenous antibiotics (cefuroxime and metronidazole) but the risk of perforation and widespread peritonitis on this regime is such that it cannot be recommended for routine use. However, in certain circumstances where surgery is impractical (e.g. on a ship without a surgeon on board), this may be the only feasible approach.

The treatment of patients presenting with an appendiceal mass is rather different. Although some surgeons favour an operative approach the majority will treat an appendix mass conservatively. If the patient has a fever and a high white count and the mass is tender these are indications that an abscess has formed. This can be confirmed by ultrasound or CT scanning and insertion of a percutaneous drain is the current treatment of choice. Occasionally this can be followed by the development of a faecal fistula but this is usually a low-output fistula which normally heals spontaneously. If percutaneous drainage is inadequate, it may be necessary to carry out operative drainage through an incision placed lateral to the mass.

If, on the other hand, a mass develops without the signs and symptoms of an abscess the best approach is conservative. However, it is important that the patient should be investigated and a barium enema about 2 months after the development of the mass is necessary to exclude other pathology, such as a caecal carcinoma. In patients who have had an appendix mass treated conservatively, about 15% will develop recurrent appendicitis. This should be explained to the patient and an interval appendicectomy can be carried out according to a decision made jointly by the patient and the surgeon.

When a patient has a laparoscopy or laparotomy which reveals an apparently normal appendix and no other pathology, there is some debate as to whether or not to carry out an appendicectomy. Given that the risk of developing appendicitis in adults is in the region of 0.04 and the risk from death from acute appendicitis is about one in 800, it seems reasonable to leave the appendix *in situ*. The reality is that when a normal appendix is seen at laparoscopy most surgeons skilled in laparoscopic appendicectomy tend to remove the appendix. Especially with the open approach most surgeons perform appendicectomy for the following reasons.

- If a scar is present in the right iliac fossa, a future assumption that appendicctomy has been carried out may be made.
- If the patient has recurring right iliac fossa pain, appendicectomy will rule out an important cause of the symptom.
- About 20% of normal-appearing appendices show microscopic evidence of mucosal ulceration and pus in the lumen.
- The majority of cases of carcinoid of the appendix occur in organs which look macroscopically normal.

Prognosis

The overall mortality rate of 0.2–0.8% from appendicitis is attributable to complications of the disease rather than to the surgical treatment. In children, the mortality rate ranges from 0.1% to 1%; in patients older than 70 years, the rate rises above 20%, primarily because of diagnostic and therapeutic delay. Appendiceal perforation is associated with increased morbidity and mortality.

Complications occur in 1–5% of patients with appendicitis, and postoperative wound infections account for almost one-third.

Neoplasms of the appendix

These are rare tumours and the vast majority present as acute appendicitis, with the diagnosis being established by histology. In this respect, it has been estimated that 1.0% of all appendicectomy specimens contain a neoplasm. The age-adjusted incidence of malignant neoplasms of the appendix is 0.12 cases per 100 000 population per year (SEER database). The gene alterations of non-carcinoid tumours of the appendix differ from those of colorectal carcinomas and consist of frequent K-*ras* proto-oncogene mutation without any alterations of *p53* tumour suppressor gene or MSI genotype. Pathologically there is an important distinction between mucinous (mucin secreting) and non-mucinous tumours of the appendix as both the clinical presentation and overall survival differ between the two groups. The five common histological types in order of descending frequency are:

- malignant carcinoid of the appendix
- GCC of the appendix
- colonic adenocarcinoma of the appendix
- mucinous adenoma/adenocarcinoma of the appendix
- signet ring carcinoma of the appendix.

Carcinoid tumours of the appendix

Carcinoids are the most common tumours of the appendix, with a peak incidence in the third and fourth decades of life. They are composed of neuroendocrine cells and exhibit variable biological behaviour from relatively benign to aggressively malignant. Carcinoid tumours are discovered in roughly one in every 300 routine appendicectomies for acute appendicitis. Occasionally carcinoid tumours secrete ACTH, and these may produce the Cushing syndrome. There is debate concerning the precise origin of carcinoid tumours: whether they arise from the Kulchitsky cells at the base of Lieberkuhn's glands or from similar cells in the lamina propria and submucosa in close association with nerves (subepithelial neuroendocrine complexes).

Carcinoids form small (70% <1.0 cm) fairly circumscribed but not encapsulated greyish white tumours, situated most commonly at the tip of the appendix (70%) giving rise to a 'bell clapper' configuration, in the body (22%) and the base (7%) of the organ.

Classical appendiceal carcinoids

Many reports on carcinoid tumours indicate that malignant carcinoids are more common in women, have a younger average age at diagnosis and, in general, a better overall survival than all other histological types of appendiceal tumours. The vast majority of classical appendiceal carcinoids are less than 1.0 cm in diameter and there has never been a reported case of metastases occurring from a tumour of this size. However, there is a clear risk of metastatic disease when a carcinoid is greater than 1.5 cm in diameter. Thus if the carcinoid exceeds 1.5 cm in diameter, a right hemicolectomy with radical removal of the ileocaecal lymph nodes is advisable. For tumours between 1.0 cm and 1.5 cm, right hemicolectomy is probably unnecessary as long as the resection margins after appendicectomy are clear. If a lesion in the appendix is encountered incidentally during an abdominal operation, the appendix should be removed and subjected to a frozen section. If this confirms a carcinoid tumour which is small (<1.0 cm), no further action is required, but if the lesion is greater than 1.5 cm, a right hemicolectomy should be performed.

Goblet cell carcinoids

GCCs account for less than 5% of tumours of the appendix and exhibit histological features of both adenocarcinoma and carcinoids and though they are more aggressive than conventional carcinoids, they exhibit considerable variation in biological aggressiveness in individual patients. The histopathological features indicative of invasive and metastatic behaviour include increased number of Paneth cells, excess mucin secretion and presence of pancreatic polypeptide.

GCC arises from a pluripotent cell (stem cell) with endodermal origin with neuroendocrine and mucinous differentiation (hence they are sometimes called amphicrine cell tumours). The component cells contain two types of granules: neuroendocrine and mucinous. The degree of mucin production varies from cells distended by mucin to discrete granules of mucin. Paneth cells are present in 30–50% of tumours and are associated with an increased risk for metastatic spread especially to the ovaries. Macroscopically, GCCs usually arise in the distal segment of the appendix (70%) from the lamina propria and then infiltrate the wall of the appendix without forming a distinct mass. The serosa and mesoappendix are involved in 20% and 50%, respectively, at the time of diagnosis. Metastases are rare when the disease does not extend beyond the muscularis propria.

Clinical features

A major review by Payam *et al.* based on 57 publications involving 600 patients indicates the mean age of presentation of GCC is 59 years, which is significantly older than the mean age of presentation of the classical carcinoids. The sex incidence is equal. The tumour is rarely diagnosed preoperatively. Its most common clinical presentation is with acute appendicitis (23%). Other manifestations include ill-defined lower abdominal pain (5%) and an appendicular mass (3%). In women, they may also present as Krukenberg tumours in the ovaries. Distant metastases are present at the time of diagnosis in 11%, with the ovaries being the most common site. Regional lymph node involvement is present in 9% of patients.

Surgical treatment

Appendicectomy is the treatment of choice in the majority of patients with no concomitant caecal involvement and low-grade tumour histology. A radical right hemicolectomy is however indicated in tumours with cellular de-differentiation, increased mitotic activity, and involvement of the base of the appendix with caecal wall infiltration, lymph node involvement and for tumour size greater than 1.5 cm. In addition to right hemicolectomy, some advocate bilateral oophorectomy in female patients because of the possibility of ovarian metastases.

Adjuvant therapy

Most of the regimes used include 5FU. Although 5FU is not used routinely by all centres, there is some evidence for its benefit in preventing peritoneal disease, and, for this reason, 5FU is recommended in all GCC patients.

Patients with residual macroscopic disease should be considered for cytoreductive surgery and intraperitoneal chemotherapy with either mitomycin C and 5FU or cisplatin and doxorubicin.

Patients with Krukenberg's tumours have been treated with cisplatin-based chemotherapy following excision with a survival benefit over surgery alone. One series with FOLFOX chemotherapy (oxaliplatin combined with 5FU and leucovorin) reported good partial and complete response with sustained remission in patients with metastasis to ovaries, peritoneal cavity and lymph nodes.

Prognosis of goblet cell carcinoid

This is estimated to be intermediate between classical carcinoids and well-differentiated adenocarcinomas of the appendix, although the outcome varies in the individual patient. Tumour characteristics that predict aggressive behaviour include size, histological subtype and mesoappendiceal involvement. Additionally, tumours larger than 2.0 cm are associated with a poorer prognosis. However, even smaller tumours do occasionally metastasize. However, the overall survival for GCC of the appendix is not significantly worse than that for patients with malignant carcinoid when adjusted for age and extent of disease at presentation. The reported average survival of GCC is 4.6 years with an extremely variable range of 6 months to 20 years.

Colonic adenocarcinoma of the appendix

Sometimes referred to as 'colonic-type adenocarcinoma of the appendix', this is similar to adenocarcinoma of the colon. It is an extremely rare tumour and accounts for approximately 15% of all malignant tumours of the appendix. When it occurs at the base of the appendix, it may be difficult to be sure whether it has arisen in the appendix or the caecum. It behaves more like colon cancer in its propensity to metastasize via the lymphatic nodal system and the bloodstream. It tends to be an aggressive tumour and 30% of these tumours have spread to the peritoneal cavity at the time of diagnosis. The disease is usually encountered in elderly people (60–70 years) and is commoner in males with a sex ratio of 2:1.

Clinical features

The majority of patients present with acute appendicitis, and in 50% of these the tumour is found to have perforated with local peritonitis. Other patients present with a hard mass in the right iliac fossa, pain and ascites. Cases with appendiceal perforation are likely to have seedling deposits and a high risk of intraperitoneal recurrence. The tumour often metastasizes to the ovaries in females, in whom an ovarian mass may be the first presentation of the disease.

Treatment

Surgical treatment consists of right hemicolectomy with or without bilateral salpingo-oophorectomy as a prophylaxis against Krukenberg tumours in females. Although there is some debate on the need for bilateral salpingo-oophorectomy, the high risk for metastatic disease in the ovaries is undeniable. Some also advocate the use of intraperitoneal chemotherapy in addition to a right hemicolectomy even in the absence of macroscopic disease within the peritoneal cavity after resection, especially for intermediate to high-grade tumours and in cases with perforation. Cytoreduction surgery and hyperthermic intraperitoneal chemotherapy are recommended for peritoneal metastasis or peritoneal carcinomatosis. Systemic chemotherapy is with regimens used for colon cancer (5FU, leucovorin, oxaliplatin, irinotecan etc.), although there have been no randomized trials to confirm a survival advantage of this chemotherapy for adenocarcinoma of the appendix. The antiangiogenic humanized monoclonal antibody Avastin has also been used but the reported experience is limited. The overall survival in the reported 5-year literature is 40–50%, but higher survival rates up to 80% have been reported in recent years even in patients with peritoneal disease after complete cytoreductive surgery and hyperthermic intraperitoneal chemotherapy.

Signet ring carcinoma of the appendix

Signet ring cell carcinomas (stomach, small bowel, colon, appendix) are composed of mucus-producing tumour cells with strong immunoreactivity against cytokeratin 20, CDX-2, MUC-2 (Mucin 2 coded by gene *Mc2*) and CEA, and focal immunopositivity for MUC-5AC (Mucin 5Ac coded by gene *Muc5a*). CDX-2 is a useful marker to confirm an appendiceal origin of pseudomyxoma peritonei (PMP). The *Cdx2* which encodes for CDX2 protein is a homeobox gene that is involved in the development of the small and large intestine and in the differentiation of intestinal epithelial cells. In adult humans, it is expressed in intestinal epithelial cells, where it acts as a transcription factor. The highest frequency of extensive CDX2 expression is seen in colorectal adenocarcinomas and mucinous appendiceal tumours. NETs of the gut also express CDX2, but the expression varies, being high in the midgut including the appendix (90%) and lower in the hindgut (44%).

Primary signet ring cell carcinoma of the appendix is exceedingly rare with an estimated incidence of 0.12 cases per 1 million people per year. Nonetheless it is considered separately from other carcinomas in view of its uniformly poor prognosis. This has been confirmed by many reported series with overall survival rates varying from 7% to 20%. As with other signet ring cell carcinomas (stomach and colon), patients with signet ring cell carcinoma of the appendix present with advanced disease that in part explains the bad prognosis. However, the signet ring cell morphology itself contributes to the poor clinical outcome even when tumour stage and disease extent are taken into account. The mean age of patients at presentation is 58 years and both sexes are affected equally. As with other tumours of the appendix, the usual presentation is with acute appendicitis and the diagnosis is established by histological examination the appendix.

Treatment

The standard surgical treatment is by right hemicolectomy. Routine bilateral salpingo-oophorectomy is advisable especially in postmenopausal females as, apart from assisting with tumour staging, it abolishes the risk of ovarian deposits which are common in signet ring tumours.

The treatment options for metastatic disease include systemic chemotherapy alone, cytoreductive surgery with peritonectomy with hyperthermic intraoperative intraperitoneal chemotherapy and combination of treatments. Of these, cytoreductive surgery and hyperthermic intraperitoneal chemotherapy have recently become the treatment of choice as for all types of mucinous-type tumours as it improves the survival rate and reduces the recurrence rate in patients with PMP. Increasingly cytoreduction and hyperthermic intraperitoneal chemotherapy is followed by systemic chemotherapy. The most important prognostic factor for survival is the completeness of cytoreduction. However, incomplete cytoreduction and perioperative intraperitoneal chemotherapy can achieve a limited long-term survival. Right hemicolectomy alone does not confer a survival advantage in patients with mucinous appendiceal tumours with peritoneal seeding and is not recommended in patients with inoperable peritoneal disease which precludes substantial cytoreduction.

Systemic chemotherapy is used on its own or after cytoreduction and intraperitoneal chemotherapy with agents used in intraperitoneal chemotherapy including mitomycin C, fluoropyrimidines and platinum compounds. Other regimens for systemic chemotherapy are based on 5FU, leucovorin and cisplatin/oxaliplatin (FOLFOX-4).

Mucinous neoplasms of the appendix

A simple mucocele of the appendix is a rare condition, which is thought to arise as a sequel to obstruction of the appendix without the onset of infection. The appendix becomes distended by mucoid secretion and the normal mucosa becomes replaced by a single layer of mucus-secreting cells. Eventually the lesion may calcify (Figure 29.33).

Mucocele of the appendix is no longer considered a benign tumour but is regarded as a low-grade borderline malignant lesion or, more appropriately, mucinous tumour of uncertain malignant potential (UMP) to distinguish it from the overtly more aggressive mucinous adenocarcinoma. In contrast to the latter, tumours in the UMP category appear benign on histology, rarely metastasize to lymph nodes and, although they may give rise to PMP, they rarely invade parenchymal organs except for the ovary or spread to regional lymph nodes. Thus PMP caused by UMP tumours (usually following rupture) carries a much better prognosis than in cases caused by mucinous adenocarcinomas.

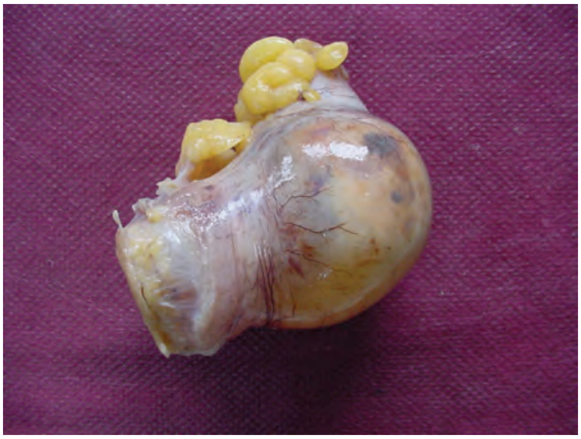


Figure 29.33 Benign mucocele of the appendix. (Courtesy of Professor Omar Shah, Srinagar, Kashmir India.)

One classification of appendiceal mucinous tumours proposed by Pai and Longacre qualifies the designation UMP for mucinous tumours of uncertain malignant potential with M-UMP for mucinous tumours which have spread to the pertinoneum but are clearly not invasive. The details of this classification are as follows.

1. *Adenoma*: simple or focally stratified columnar epithelium with goblet cells; mild to moderate atypia; no atypical mitosis, no stromal invasion, no extra-appendiceal epithelium – its oncological significance is that when completely excised without rupture it does not recur; but rupture may cause acellular mucinous ascites.
2. *Mucinous tumour of UMP*: as in 1 but (1) with involvement of proximal margin, (2) mucin with epithelium though present within the wall is not overtly invasive, (3) there is uncertainty on the presence of epithelium within extra-appendiceal mucin – its oncological significance relates to the risk of PMP and need for close follow-up.
3. *Mucinous tumour of M-LMP*: as in 1, but neoplastic cells are present in peritoneal implants. May require hyperthermic intraperitoneal chemotherapy.
4. *Mucinous adenocarcinoma:* a frankly invasive tumour which may cause aggressive PMP with visceral invasion and nodal metastases.

However, there are other classifications and there seems to be a lack of agreement amongst expert pathologists in defining a generally accepted classification and some have suggested a simpler histological grouping into (1) appendiceal mucinous neoplasm with low dysplasia, (2) appendiceal mucinous neoplasm with high-grade dysplasia and (3) invasive mucinous adenocarcinoma.

Clinical features

Mucinous tumours often present with acute appendicitis. Other presentations include abdominal distension from mucinous collection in the abdomen or pelvis or more generalized ‘mucinous ascites’. Females may present with ovarian masses

(Krukenberg tumours). Male patients may present with inguinal hernia caused by the increased intra-abdominal pressure.

Treatment

A 'benign' mucinous adenoma that has not ruptured and which is confined to the appendix is treated by appendicectomy with extreme care being taken to avoid rupture during removal especially if the laparoscopic approach is used when retrieval should be through a large enough wound to ensure against this complication.

Treatment for suspected mucinous adenocarcinoma confined to the appendix should be by *en bloc* removal of the appendix with the right colon. Systemic chemotherapy is used in addition for treatment of mucinous adenocarcinoma.

Mucin genes, which play an important role in the pathogenesis of PMP, are regulated by several mucin-encoding genes and by epidermal growth factor receptor. This has recently led to treatment of PMP with cetuximab monotherapy. PMP is considered in Chapter 21.

Meckel's diverticulum

Anatomy

A Meckel's diverticulum is a remnant of the vitellointestinal duct and is present in about 2% of the population. It arises from the antimesenteric side of the ileum. It has the same microscopic structure as the adjacent small bowel and it has a separate blood supply from the adjacent SBM (the omphalomesenteric artery). Although it is frequently said that most Meckel's diverticula are situated two feet (60 cm) from the ileocaecal valve, as many as 25% may be situated more proximally in the ileum. The length and shape of the diverticulum is highly variable and although 85% are blind-ended, the rest have an attachment which is related to its embryological origin. Two per cent of these exist as a patent vitellointestinal duct with a faecal fistula at the umbilicus. In 1%, the diverticulum is attached to the umbilicus by a band, and in about 10% of cases a band from the diverticulum is connected to the adjacent mesentery (a mesodiverticular band) (Figure 29.34).

Ectopic tissue is found within the Meckel's diverticulum in about 70% of cases. This is usually gastric mucosa, but it is also possible to find pancreatic, duodenal or colonic tissue. The presence of ectopic gastric mucosa has important implications for the clinical presentation of Meckel's diverticulum, as it is associated with peptic ulceration secondary to acid and pepsin secretion.

Pathology

The majority of Meckel's diverticula do not cause problems and remain asymptomatic. However, the following three pathological processes may occur as a result of a Meckel's diverticulum:

- inflammation of the diverticulum
- peptic ulceration of the small bowel
- intestinal obstruction.

Inflammation

Meckel's diverticulum can become inflamed in very much the same way as the vermiform appendix. Likewise, gangrene and perforation of the diverticulum may ensue.

Peptic ulceration

When there is heterotopic gastric mucosa within the lumen of the diverticulum (~40%) this secretes pepsin and hydrochloric acid. As a result peptic ulceration may occur, most commonly at the neck of the diverticulum or just distally in the ileum. The role of H. *pylori* in peptic ulceration related to a Meckel's diverticulum is not known.

Intestinal obstruction

Intestinal obstruction may result from intussusception occurring with the diverticulum acting as a lead point. It can also occur because of persistence of the band which was once the vitellointestinal duct producing a small bowel volvulus. In addition, entrapment of small bowel can occur through a defect caused by a mesodiverticular band or an omphalomesenteric artery.

Clinical features

The clinical presentation of asymptomatic Meckel's diverticulum depends on the underlying pathology.

Acute inflammation

Acute Meckel's diverticulitis produces a symptom complex which is very similar to appendicitis but normally the pain persists in the central abdominal area without a shift to the right lower quadrant. This usually occurs in children and frequently a diagnosis of acute appendicitis is made.

Peptic ulceration

Peptic ulceration of the neck of the Meckel's diverticulum or the ileum can lead to bleeding which presents either as melaena or fresh rectal bleeding. Again this nearly always occurs in children.

Intestinal obstruction

This may present with the typical clinical features of a small bowel obstruction. However, infarction of the bowel is likely with volvulus or entrapment of the small bowel by the band and in such cases the patient will become rapidly ill with a pyrexia and tachycardia, and be found to have peritonitis on examination.

Diagnosis

The diagnosis of Meckel's diverticulum is usually made at operation but in the child who is having repeated episodes of brisk rectal bleeding or melaena, a technetium scan may be of value. In this test, intravenous administration of ^{99m}Tc is followed by external scintiscanning. The radionuclide is taken up by the ectopic gastric mucosa. Mesenteric angiography may also be helpful in the presence of active bleeding.

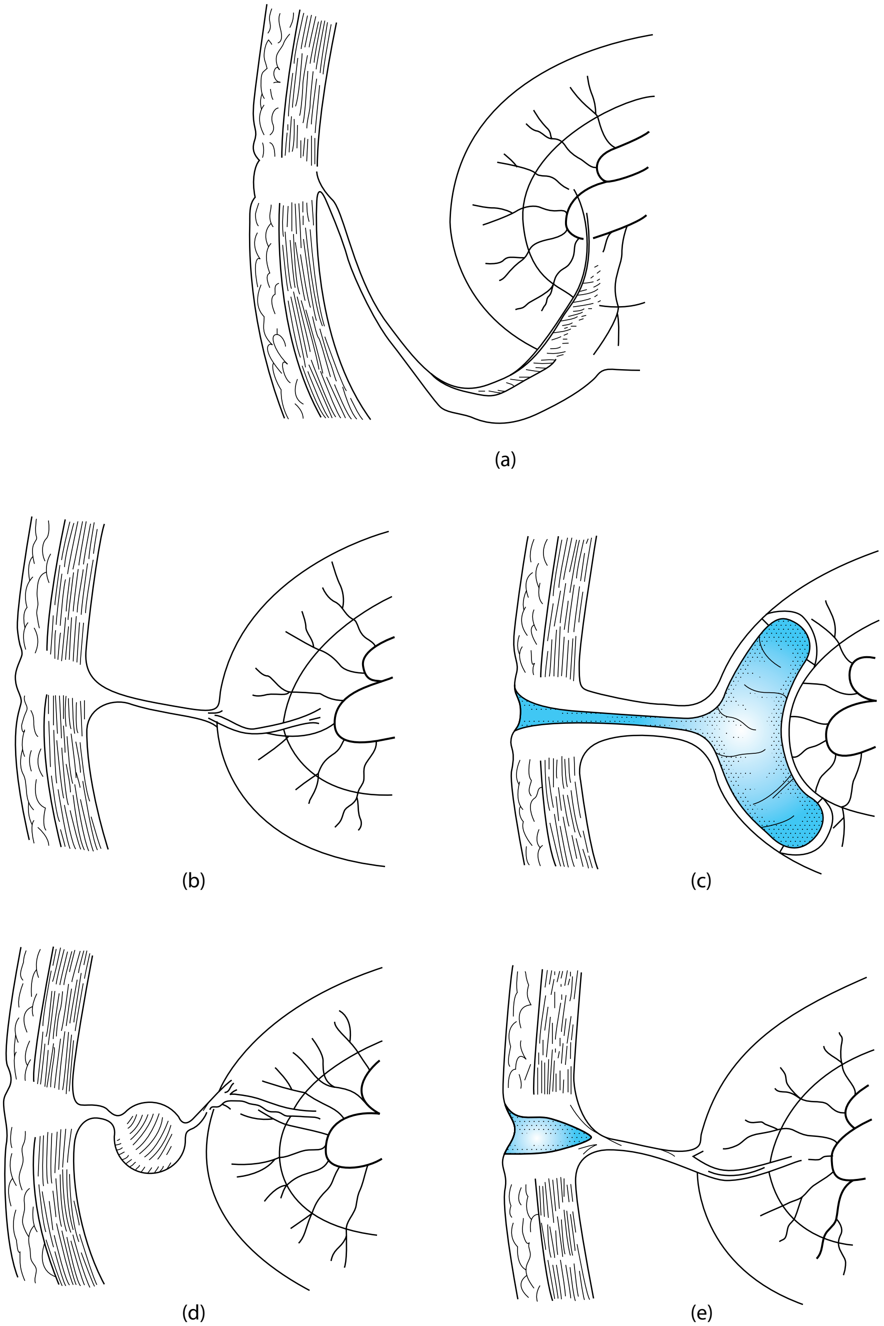


Figure 29.34 The vitellointestinal duct may connect with the umbilicus as (a) a fibrous cord extending from a Meckel's diverticulum; (b) a simple fibrous cord to a loop of ileum; (c) as an umbilical fistula; (d) a fibrous cord with a cyst; (e) an umbilical sinus.

Treatment

The standard treatment of a Meckel's diverticulum is excision of the diverticulum together with a wedge of the adjacent ileum. In some cases, however, with extensive inflammation of the diverticulum a limited small bowel resection may be necessary. If a Meckel's diverticulum is found incidentally during the course of a laparotomy/laparoscopy, removal is not necessary if the diverticulum has a wide base and feels soft on palpation. However, if the neck is narrow or nodules are palpable in the walls of the diverticulum, removal is advisable.

GUIDE TO FURTHER READING

Al-Batran SE, Hartmann JT, Heidel F, *et al.* Focal progression in patients with gastrointestinal stromal tumors after initial response to imatinib mesylate: a three-center-based study of 38 patients. *Gastric Cancer* 2007;**10**:145–152.

Binion DG. Biologic therapies for Crohn's disease: Update from the 2009 ACG Meeting. *Gastroenterol Hepatol* 2010;**6**(Suppl. 2):4–16.

Caprilli R, Frieri G. The dyspeptic macrophage 30 years later: An update in the pathogenesis of Crohn's disease. *Digest Liver Dis* 2009;**41**:166–168.

Chadalavada R, Brown DK, Walker AN, Sedghi S. Cronkhite-Canada syndrome: sustained remission after corticosteroid treatment. *Am J Gastroenterol* 2003;**98**:1444–1446.

Chang HK, Yu E, Kim J, *et al.* The Korean Small Intestinal Cancer Study Group. Adenocarcinoma of the small intestine: a multi-institutional study of 197 surgically resected cases. *Hum Pathol* 2010;**41**:1087–1096.

d'Amore F, Brincker H, Grønbaek K, *et al.* Non-Hodgkin's lymphoma of the gastrointestinal tract: a population-based analysis of incidence, geographic distribution, clinicopathologic presentation features, and prognosis. Danish Lymphoma Study Group. *J Am Soc Clin Oncol* 1994;**12**:1673–1684.

De Angelis P, Morino G, Pane A, *et al.* Eosinophilic esophagitis: management and pharmacotherapy. *Expert Opin Pharmacother* 2008;**9**:731–740.

Dignass A, Van Assche G, Lindsay JO, *et al.* The Second European Evidence-Based Consensus on the Diagnosis and Management of Crohn's Disease: current management. *J Crohn's Colitis* 2010;**4**:28–62.

Ferreri AJ, Zinzani PL, Govi S, Pileri SA. Enteropathy-associated T-cell lymphoma. *Crit Rev Oncol/Hematol* 2011;**79**:84–90.

Givi B, Pommier SJ, Thompson AK, *et al.* Operative resection of primary carcinoid neoplasms in patients with liver metastases yields significantly better survival. *Surgery* 2006;**140**:891–897.

Hyperlink-Fletcher C, Berman J, Corless C, *et al.* Diagnosis of gastrointestinal stromal tumors: a consensus approach. *Hum Pathol* 2002;**33**:478–483.

Iwanaka T, Arai M, Ito M, *et al.* Surgical treatment for abdominal neuroblastoma in the laparoscopic era. *Surg Endosc* 2001;**15**:751–754.

Junquera F, Quiroga S, Saperas E, *et al.* Accuracy of helical computed tomographic angiography for the diagnosis of colonic angiodysplasia. *Gastroenterology* 2000;**119**:293.

Junquera F, Saperas E, Videla S, *et al.* Long-term efficacy of octreotide in the prevention of recurrent bleeding from gastrointestinal angiodysplasia. *Am J Gastroenterol* 2007;**102**:254.

Levy AD, Abbott RM, Rohrmann CA Jr, *et al.* Gastrointestinal hemangiomas: imaging findings with pathologic correlation in pediatric and adult patients. *AJR Am J Roentgenol* 2001;**177**:1073–1081.

Lichtenstein GR, Hanauer SB, Sandborn WJ, *et al.* Management of Crohn's disease in adults. *Am J Gastroenterol* 2009;**104**:465–483.

Malek SN, Hatfield AJ, Flinn IW. MALT lymphomas. *Curr Treat Options Oncol* 2003;**4**:269–279.

Monkemuller K, Neumann H, Evert M. Cronkhite-Canada syndrome. Panendoscopic characterization with esophagogastroduodenoscopy, endoscopic ultrasound, colonoscopy, and double balloon enteroscopy. *Clin Gastroenterol Hepatol* 2008;**6**:A26.

Moscandrew M, Mahadevan U, Kane S. General health maintenance in IBD. *Inflammatory Bowel Dis* 2009;**15**:1399–1409.

Noomen, CG, Hommes DW, Fidder HH. Update on genetics in inflammatory disease. *Best Pract Res Clin Gastroenterol* 2009;**23**:233–243.

Novitsky YW, Kercher KW, Sing RF, Heniford BT. Long-term outcomes of laparoscopic resection of gastric gastrointestinal stromal tumors. *Ann Surg* 2006;**243**:745–747.

Pahlavan PS, Kanthan R. Goblet cell carcinoid of the appendix. *World J Surg Oncol* 2005;**3**:36.

Pai RK, Longacre TA. Appendiceal mucinous tumours and pseudomyxoma peritonei: histologic features, diagnostic problems, and proposed classification. *Adv Anat Patholol* 2005;**12**:291–311.

Raju GS, Gerson L, Das A, *et al.* American Gastroenterological Association (AGA) Institute technical review on obscure gastrointestinal bleeding. *Gastroenterology* 2007;**133**:1697.

Rha SE, Byun JB, Jung SE, *et al.* Neurogenic tumors in the abdomen: tumour types and imaging characteristics. *Radio graphics* 2003;**23**:29–43.

Salomon L, Rabii R, Soulie M, *et al.* Experience with retroperitoneal laparoscopic adrenalectomy for pheochromocytoma. *J Urol* 2001;**165**:1871–1874.

Samoha S, Arber N. Cronkhite-Canada syndrome. *Digestion* 2005;**71**:199–200.

Sandborn WJ, Feagan BG, Stoinov F, *et al.* Certolizumab pegol for the treatment of Crohn's disease. *N Engl J Med* 2007;**357**:228–238.

van Oosterom AT, Judson IR, Verweij J, *et al.* Update of phase I study of imatinib (STI571) in advanced soft tissue sarcomas and gastrointestinal stromal tumors: a report of the EORTC Soft Tissue and Bone Sarcoma Group. *Eur J Cancer* 2002;**38**(Suppl. 5):S83–87.

Wu PC, Langerman A, Ryan CW, *et al.* Surgical treatment of gastrointestinal stromal tumors in the imatinib (STI-571) era. *Surgery* 2003;**134**:656–665; discussion 665–666.

Yano T, Yamamoto H, Sunada K, *et al.* Endoscopic classification of vascular lesions of the small intestine (with videos). *Gastrointest Endosc* 2008;**67**:169.

CHAPTER 30

Disorders of the colon and rectum

PARIS TEKKIS, PANAYIOTIS GEORGIOU, EDWARD FITZGERALD AND EMIL TAN

Anatomy

The large intestine consists of the caecum, vermiform appendix, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, rectum and anal canal (Figure 30.1). The anatomy of the appendix and anal canal is dealt with separately in the appropriate chapters.

Originally a mid-line structure, the large intestine undergoes rotation during embryological development and as a result the ascending colon and the descending colon are essentially retroperitoneal structures. However, the degree to which the large intestine has a mesentery is highly variable as is its total length that averages 1.5 m. The whole of the large intestine is capable of considerable distension, although in an adult living in the Western world the left side of the colon tends to be less distensible than the right owing to muscular hypertrophy.

The caecum lies in the right iliac fossa and is approximately 7 cm in length and width. Proximally it becomes the ascending colon at its junction with the terminal ileum. The caecum lies on the iliac and psoas muscles and on the genitofemoral, femoral and lateral cutaneous nerves. It also lies anterior to the testicular or ovarian vessels and the ureter. The exact position of the caecum is variable and it may extend into the true pelvis. The caecum is almost completely surrounded by peritoneum but it is often attached to the iliac fossa medially and laterally. This can produce a retrocaecal peritoneal recess which may extend upwards posterior to the ascending colon.

The ileocaecal junction is extremely variable in appearance. In most circumstances the ileum enters obliquely into the large bowel through a horizontal slit and is partly invaginated into the caecum to form a fold (the ileocaecal valve). Reflux of caecal contents into the small bowel is prevented by contraction of the circular muscle of the ileum which leads to closure of the ileocaecal valve. However, the muscle in the valve is poorly developed and the ileocaecal valve is frequently incompetent.

The ascending colon varies from about 10 to 20 cm in length. It lies on the iliac muscle, the iliac crest and quadratus lumborum, crossing the lateral cutaneous nerve of the thigh, the ilioinguinal and iliohypogastric nerves. It ends at the hepatic flexure where the large bowel turns to the left on the lower portion of the right kidney inferior to the liver. Under most circumstances peritoneum covers the front and the sides of the ascending colon and fixes it firmly to the posterior abdominal wall, but occasionally there is a short mesentery.

The transverse colon is the longest section of the colon varying from 40 to 70 cm in length. It extends from the hepatic flexure to the splenic flexure and forms a dependent loop between both of these points. The lowest point of the transverse colon may reach below the umbilicus, although it is usually just superior to it. The transverse colon is suspended by the transverse mesocolon which is fused to the posterior surface of the greater omentum. This transverse mesocolon is attached to the descending part of the duodenum, to the head and the lower aspect of the body of the pancreas and to the anterior surface of the left kidney. It contains both the middle colic vessels and branches of the right and left colic

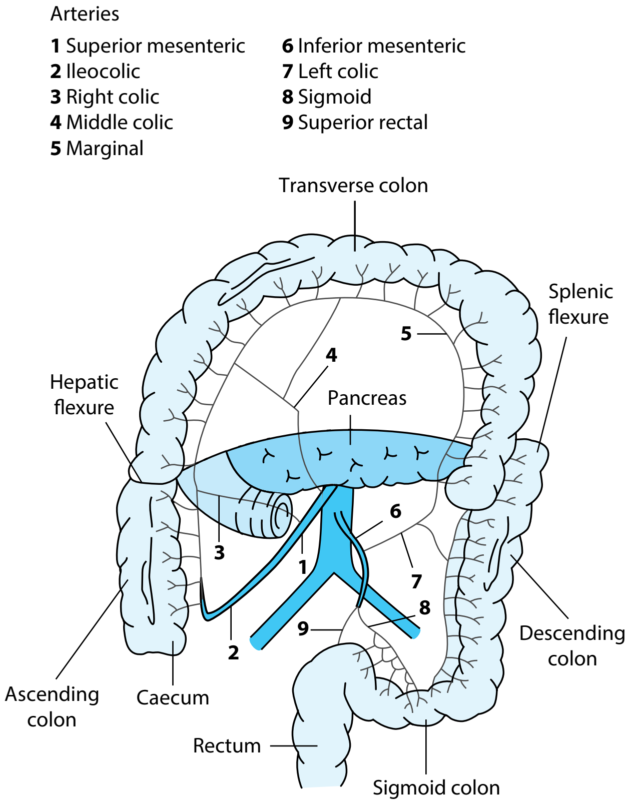


Figure 30.1 The large intestine and its blood supply.

vessels with associated nerves and lymphatics. Thus the transverse colon starts immediately anterior to the descending part of the duodenum and the head of the pancreas, descends anterior to the small intestine and ascends to the splenic flexure. At this point it is anterior to the left part of the left kidney and immediately below the spleen. The splenic flexure is attached to the diaphragm by peritoneum (phrenicocolic ligament) and can be extremely close to the spleen. At this point the greater omentum frequently has attachments to the spleen and is closely associated with the colon. Traction on these splenic attachments may cause splenic bleeding in the course of mobilization of the splenic flexure.

The descending colon extends from the splenic flexure to the rim of the true pelvis close to the inguinal ligament. The descending colon is attached by peritoneum to the posterior abdominal wall in the left paravertebral gutter and iliac fossa. Superiorly it is anterior to the lateral surface of the left kidney and medial to the diaphragm, and then it lies on the same muscles and nerves as the ascending colon. At the anterior superior iliac spine, the descending colon turns medially, superior to the inguinal ligament, and lies on the femoral nerve, psoas muscle and the genital vessels, and becomes the sigmoid colon immediately anterior to the external iliac vessels.

The sigmoid colon is the most variable part of the colon in terms of its length (50–80 cm) and its mobility. It extends from the end of the descending colon to the rim of the true pelvis where it becomes the rectum. It has a long mesentery (the sigmoid mesocolon) which has quite a short base starting at the end of the descending colon and ascending on the external iliac vessels to the mid - point of the common iliac artery. At this point it turns downwards and to the right to the rim of the true pelvis. The mesocolon contains the sigmoid branches of the inferior mesenteric artery and associated nerves and lymphatics. Under normal circumstances the sigmoid colon lies entirely in the left iliac fossa and the true pelvis but it may also extend across to the right iliac fossa.

The rectum is that part of the large bowel which lies in the true pelvis at the point where the sigmoid mesocolon ends. Again, this is highly variable in length depending very much on the build of the individual but it is said to be 15 cm long as measured by a rigid sigmoidoscope. It follows the curve of the sacrum and the coccyx and then runs anteriorly and inferiorly to the central perineal tendon lying on the anococcygeal ligament and the levator ani muscles. It then ends by turning posteriorly and inferiorly as the anal canal, immediately posterior to the central perineal tendon and to the apex of the prostate in the male. The lowest part of the rectum is more capacious than the rest and is known as the ampulla. The rectum is not straight; in the sagittal plane it follows the curve of the sacrum and coccyx and in the coronal plane it is S shaped. This gives rise to prominent folds within the lumen of the rectum known as the valves of Houston. The front and the sides of the upper third of the rectum are covered with peritoneum but this gradually moves anteriorly and turns off the front of the rectum at the junction between its middle and lower thirds. This forms the rectouterine or rectovesical pouch by passing upwards on the back of the posterior fornix of the vagina or the back of the bladder respectively in the female and the male. In its lower third, the rectum lies behind the base of the bladder, the seminal vesicles and the prostate in the male and behind the vagina in the female.

In both sexes the rectum and its surrounding areolar tissue is separated from the anterior structures by a fascial layer known as Denonvilliers' fascia. Posteriorly the rectum is separated from the sacrum and the coccyx and anococcygeal ligament and the muscles attached to these (piriformis and levator ani) by a layer of pelvic fascia. This fascia is known as Waldeyer's fascia. In its upper two-thirds the actual muscular wall of the rectum is separated from the pelvic fascia by a posterior cushion of areolar tissue which becomes circumferential below the rectouterine or rectovesical pouch. This carries the blood supply to the rectum and its lymphatic drainage and is known as the mesorectum (Figure 30.2). Inferiorly and posteriorly the mesorectum has a bilobed structure.

Taeniae coli

The taeniae coli are three ribbon-like thickenings of the otherwise thin longitudinal muscle of the large bowel which arise from the longitudinal muscle at the root of the vermiform appendix and end by spreading out at the end of the sigmoid colon to become continuous with the thicker longitudinal muscle of the rectum. These three taeniae are spaced out uniformly around the circumference of the colon and between them the wall of the colon bulges outwards forming pouches or sacculations. In the ascending colon and descending colon, the taeniae are anterior, posteromedial and posterolateral, whereas in the transverse colon the positions become posterior, superior and anterior.

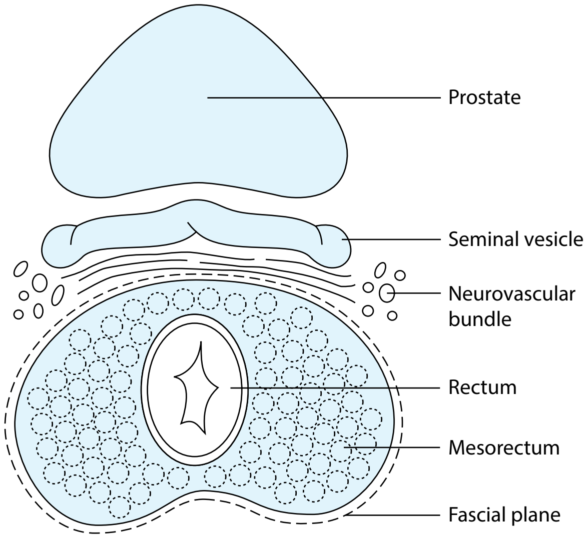


Figure 30.2 Cross-section of the rectum at the level of the seminal vesicles showing the mesorectum.

Blood supply of the large intestine

The most important vessels involved in the blood supply of the colon are the superior mesenteric artery, the inferior mesenteric artery and the marginal artery, which supplies an anastomosis between these two vessels (Figure 30.1). The superior mesenteric artery originates from the aorta just below the coeliac axis and passes posterior to the pancreas. Its terminal branch which supplies the caecum is known as the ileocolic artery. Its other named main branches which supply the colon are the right colic artery, which supplies the ascending colon, and the middle colic artery, which runs in the transverse mesocolon to supply the transverse colon. The inferior mesenteric artery arises from the aorta just inferior to the third part of the duodenum and descends on the left side of the aorta posterior to the peritoneum. Its first branch is the left colic artery which passes to the left and divides into ascending and descending branches. The ascending branch supplies the left side of the transverse colon and the splenic flexure, whereas the descending branch supplies the descending colon. The inferior mesenteric artery terminates at the base of the sigmoid mesocolon where it divides into sigmoid branches supplying the sigmoid colon and the superior rectal artery, which descends in the mesorectum to supply the upper part of the rectum. The marginal artery runs along the ascending, transverse and descending colons, receiving branches from the other colic arteries so that under normal circumstances ligation of any one of the main colic arteries would not result in ischaemia of any part of the colon. It must be remembered, however, that the marginal artery can become quite tenuous at the splenic flexure and continuity of the blood supply may be maintained by the ascending and descending branches of the left colic artery. It is therefore imperative that these two vessels are preserved when dividing either the left colic artery or the inferior mesenteric trunk during mobilization of the colon.

As mentioned above, the upper part of the rectum derives its blood supply from the superior rectal artery which is a terminal branch of the inferior mesenteric artery, but the lower rectum receives blood from the two middle rectal (or haemorrhoidal) arteries coming from the internal iliac arteries and two inferior rectal arteries which originate from the internal pudendal arteries in the ischiorectal fossae. The internal pudendal artery is itself a branch of the internal iliac artery.

The venous drainage of the large intestine follows its arterial blood supply but of course empties into the portal venous system. The inferior mesenteric vein diverges from the inferior mesenteric artery and passes up behind the pancreas to join the splenic vein. The superior mesenteric vein lies to the right of the superior mesenteric artery and joins the splenic vein at its junction with the portal vein behind the neck of the pancreas.

The lymphatic drainage of the large intestine also follows the blood supply. Small lymph nodes lie close to the marginal artery and also along the arteries leading towards it. The lymph draining through the lymph nodes associated with the branches of the superior mesenteric arteries passes into the intestinal trunk which lies in the root of the small bowel mesentery. The lymph nodes associated with the inferior mesenteric artery drain into the lumbar lymph nodes beside the aorta. Both of these empty into the cisterna chyli which enters the posterior thorax via the diaphragmatic hiatus.

Nerve supply of the large intestine

The nerve supply of the colon and the rectum, like the lymphatics, follows the course of the main vessels. The right colon receives sympathetic nerve fibres from the lower six dorsal ganglia via the superior mesenteric plexus and parasympathetic fibres from the coeliac branch of the posterior vagus nerve. The left colon and rectum are supplied through the upper three lumbar ganglia via the inferior mesenteric, superior hypogastric and inferior hypogastric plexuses (Figure 30.3). The latter plexus also receives branches from the sacral parasympathetic nerves (nervi erigentes). These nerves remain outside the pelvic fascia and are sometimes injured during mobilization of the rectum.

Physiology of the colon and rectum

Absorption acid excretion

About 1000 mL of ileal contents containing 90% water are discharged into the caecum every day in the normal adult. Water absorption takes place during transit through the colon, and only 100–200 mL of water is excreted in the faeces. The absorptive capacity of the colon depends on the rate at which ileal contents enter the caecum, and is greater in the right colon than in the left. Normal faeces are composed of 70% water and 30% solids; about 50% of the solids are bacteria and the remainder is composed of food waste and desquamated epithelium.

Nutrients such as glucose, amino acids, fatty acids and vitamins can be absorbed slowly through the colonic wall, but only very small amounts of these substances actually reach the caecum under normal circumstances. Sodium absorption is very efficient and is maintained by an active transport mechanism enhanced by mineralocorticoids and glucocorticoids. A normal adult can remain in sodium balance with as little as 5 mmol of

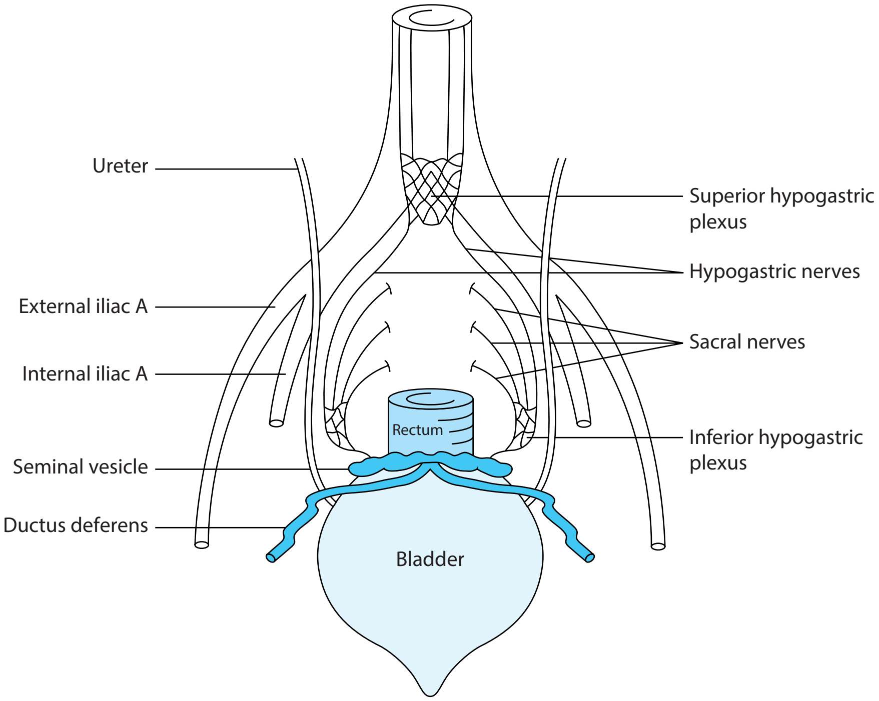


Figure 30.3 The autonomic nerves of the pelvis.

sodium per day in the diet, but following total colectomy and ileostomy the minimum daily requirement increases to about 100 mmol to offset losses from the stoma. Chloride and water absorption is passive and follows electrical and osmotic gradients established by the sodium pump.

Potassium is actively excreted into the faeces against a concentration gradient and by secretion in mucus. Excessive mucus production (e.g. in colitis or in villous adenomas) may lead to enormous losses of potassium. Only a small amount of bicarbonate is secreted into the colonic lumen in exchange for chloride.

A normal bowel habit is hard to define since it is influenced by social and dietary customs. The frequency of bowel movement in Western countries ranges from every 8 hours to once every 2–3 days. Any persistent change in bowel habit is an indication for investigation to exclude organic disease. Diarrhoea may be defined as stools containing more than 300 mL of fluid daily, although again this is highly variable. When excessive, it may be debilitating and even fatal if associated with large losses of fluid and electrolytes which are not replaced. Inflammatory disease of the colon or small bowel mucosa may cause excessive exudation of fluid and also lead to diarrhoea, as does anything that decreases the intestinal transit time and decreases absorptive surface area. The symptom of constipation has different meanings for different individuals. It may imply infrequency of bowel movements, hard consistence of the stools or difficulty in evacuation. The pathologies leading to diarrhoea and constipation are dealt with later in the chapter.

Colonic motility

There are three patterns of motor activity in the colon:

- *Segmentation.* This is the most common type of motor activity seen in the transverse and descending colon and consists of annular contractions that divide the lumen into segments propelling faeces over short distances in both directions. Segmental contractions form, relax and re-form in different locations in a random fashion, three to eight times per minute.
- *Mass movements.* These consist of strong contractions moving distally over relatively short distances (30–45 cm) in the transverse and descending colon. These are infrequent and probably occur only a few times each day, often in response to a meal.
- *Retrograde peristalsis.* This consists of annular contractions moving proximally in the right colon and in the sigmoid and descending colon. This is more frequently seen in experimental animals than in man. The retrograde movement can be shown by observing the spontaneous movement of a radio-opaque marker from the left to right colon.

A complex neurohormonal mechanism is involved in the colonic response to eating which has been inaccurately described as the gastrocolic reflex. This response consists of increased ileal emptying, increased mass movements and an urge to defecate. Other factors influencing colonic motility are physical activity, emotional states and faecal bulk. Thus normal colonic emptying is slow, complex and exceedingly variable. It is difficult to define altered motility in diseased states. There is no orderly laminar flow; some of the material entering the caecum flows past faecal material which has remained from earlier time periods. In general, residue from a meal reaches the caecum after 4 hours and the rectosigmoid by 24 hours. Since there is a large amount of mixing of bowel contents in the colon, residue from a single meal may be passed in the stool for 3–4 days afterwards under normal circumstances.

Intraluminal pressure studies of the colon can be performed by the use of small balloons, fine open-ended catheters or telemetry capsules. Such studies indicate that, although pressures of up to 100 mmHg can be generated, faecal transport can take place without a rise in intraluminal pressure. However, because specific patterns cannot be correlated with defined diseased states these investigations have little clinical significance. Rhythmic changes in the electrical potential in colonic muscle occur normally at two frequencies, three and nine per minute, respectively. The frequency appears to be approximately 16 per minute in the sigmoid colon with diverticular disease.

Microbiology of the large intestine

The normal faecal flora exists in a symbiotic relationship with the human host and supports several physiological processes. Bile pigments are degraded by colonic bacteria to give the stool its brown colour. Characteristic faecal odour is due to the amines indole and skatole, which are produced by bacterial action. Colonic bacteria also supply vitamin K to the host, alter both colonic motility and absorption and may be important in the defence against potentially more pathogenic organisms. Faecal bacteria deconjugate bile salts to produce free bile acids and also alter the steroid nucleus. These bacteria have been implicated in the pathophysiology of a variety of disease processes including the pathogenesis of carcinoma of the large bowel.

The colon of the fetus is sterile and bacterial colonization occurs soon after birth. The bacterial flora present in the colon varies with dietary and environmental factors, but over 99% of the normal flora is anaerobic. *Bacteroides fragilis* is the most prevalent. *Lactobacillus bifidus*, clostridia of various types and cocci of various types form the other common anaerobes. Aerobic bacteria can be divided into two groups, coliforms and enterococci. *Escherichia coli* is the predominant coliform and is present in counts of around 10^7 per gram of wet faeces. *Streptococcus faecalis* is the principal enterococcus and is present in similar numbers.

The bacterial flora of the colon is readily altered by antibiotic administration. Oral neomycin and tetracycline result in resistant R-factor enterococci and resistant staphylococci and bacteroides. Outbreaks of staphylococcal enterocolitis and more commonly pseudomembranous colitis from *Clostridium difficile* are frequently seen in surgical units.

Investigations

The main techniques used in the investigation of colorectal disease (excluding anorectal problems) can be subdivided into endoscopy and radiology. Under these headings the following procedures will be considered: rigid endoscopy, flexible endoscopy, barium enema, radio-opaque marker studies, colonic scintigraphy, ultrasound, CT scanning and MRI.

Rigid endoscopy

Rigid endoscopy is only of value in examining the anal canal and rectum and in the UK the nomenclature for describing the appropriate instruments is confusing. The term proctoscope refers to a short instrument which is only really of value in examining the anal canal, and this will be described in the appropriate section. The term rigid sigmoidoscope is the term used to describe the 25 cm long instrument which is used in association with a fibreoptic light source and insufflating bellows (Figure 30.4). Although reusable metal instruments are available most outpatient departments now use disposable plastic instruments which can be self-lubricating. This allows multiple examinations to be performed without the risk of infection and the need for cleansing and sterilizing the instrument. In addition, the attachment between the instrument and the bellows should be protected by a filtering device which will prevent infective agents being harboured in the bellows.

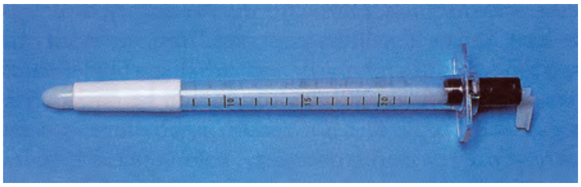


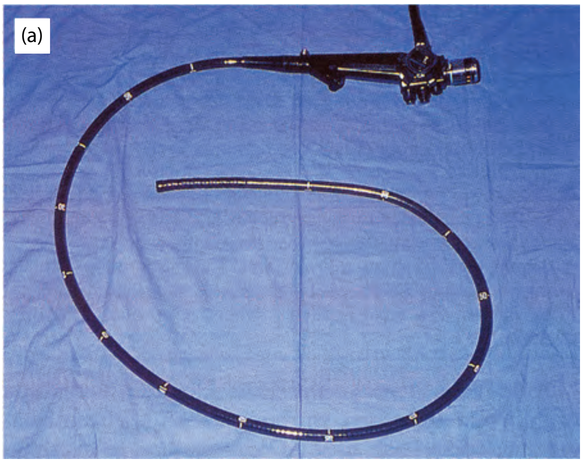
Figure 30.4 Rigid sigmoidoscope.

The term sigmoidoscope is misleading as only the very distal part of the sigmoid can be examined by the instrument and in most cases it is impractical to pass the instrument much beyond 15 cm. Under most circumstances the rigid instrument is used in unprepared bowel; although this may make the investigation difficult and sometimes impossible, it is useful to be able to see whether there is any blood staining on the faecal material. The other major advantage of rigid sigmoidoscopy is that it allows the operator to determine the exact position of a rectal lesion, i.e. whether it is on the anterior, posterior, right or left lateral walls of the rectum. This is essential information when considering transanal excision of a rectal tumour as the positioning of the patient on the operating table will depend on the exact site of the tumour. Aside from this, rigid sigmoidoscopy has no advantages over flexible endoscopy and it has several disadvantages. First, it is more uncomfortable; second, examination is limited to the rectum; third, it is difficult to see behind mucosal folds; and, fourth, it can be very difficult to see lesions just inside the anal canal.

Nevertheless, the instrument is widely used in outpatient departments as it offers the surgeon a rapid method of examining the rectum to look for tumours or mucosal inflammation. When carrying out an examination it is important to bear in mind the slightly anterior direction of the anal canal and then the sharp posterior angulation of the rectum, followed by the curve of the sacrum. Whenever pain is encountered the examination should be terminated immediately. When biopsies are taken using a rigid sigmoidoscope it must be remembered that the biopsy forceps are much larger than those used with flexible instruments, and care must be taken not to cause full - thickness damage to the rectal wall, particularly anteriorly where a perforation may occur.

Flexible endoscopy

The two types of flexible endoscope available are the sigmoidoscope and the colonoscope. The flexible sigmoidoscope is a 60 cm instrument, which is designed to examine no further



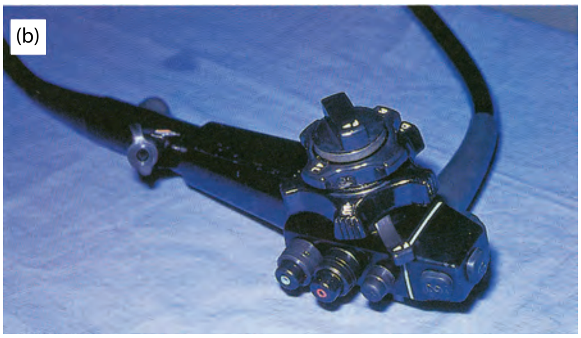


Figure 30.5 (a) A fibreoptic colonoscope. (b) The operating head of a video colonoscope. Note the absence of an eyepiece.

than the descending colon. The colonoscope (Figure 30.5), on the other hand, is a longer instrument, varying from 120 to 180 cm in length and is designed to reach the caecum. Flexible endoscopes are based on fibreoptic technology, and more recently videoendoscopes which incorporate a chip camera in the distal end of the instrument have come to the fore. In the latter, a digital image is displayed on a video monitor without an eyepiece optical interface.

The technique of flexible sigmoidoscopy is essentially similar to that of the first part of colonoscopy and therefore separate descriptions of the two procedures are not necessary. However, it is conventional for flexible sigmoidoscopy to be done without sedation whereas most endoscopists will use some form of sedation for colonoscopy as both insufflation of the large bowel and looping of the colonoscope can cause considerable discomfort. It is however essential that the patient should be able to respond to pain as this is an important signal to the endoscopist that the procedure may be about to cause damage to the colon.

Although there are many different forms of sedation it is the authors' preference to use an intravenous combination of an opiate (pethidine) and a benzodiazepine (Diazemuls). The sedation produced is profound and it is essential that secure venous access is obtained and the patient's oxygen saturation is continuously measured by pulse oximetry. It is also essential that antagonists to both opiates (nalaxone) and benzodiazepines (flumazenil) are readily available.

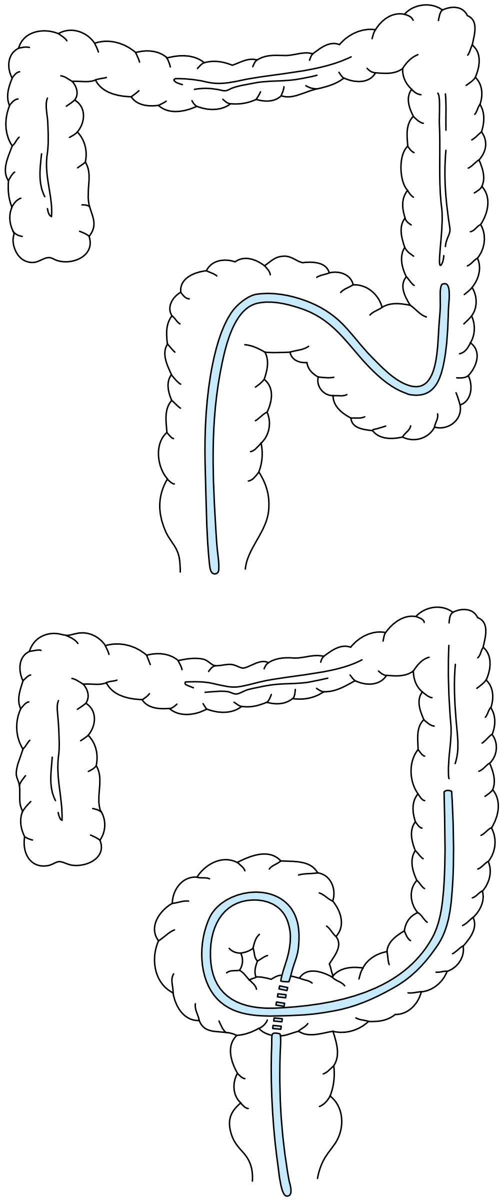


Figure 30.6 The N loop and the α loop.

The examination begins conventionally with the patient lying in the left lateral position. The rectum is usually easy to negotiate but at the rectosigmoid junction it may be necessary to pass the tip of the colonoscope 'blindly' to gain access to the sigmoid colon. This is a hazardous manoeuvre, particularly in the patient who has had previous pelvic surgery leading to adhesions, as perforation of the colon can occur. It must therefore be carried out with the utmost gentleness. In the sigmoid colon the colonoscope may assume either an N loop or an α loop (Figure 30.6). Once a loop like this has formed it is necessary to straighten the endoscope by pulling back and applying clockwise torque (Figure 30.7). The colonoscope should then be advanced maintaining the torque. If clockwise torque does not work, it is worth trying anticlockwise torque.

It is then usually a straightforward procedure to reach the splenic flexure but advancing the endoscope along the transverse colon may be difficult because of the formation of a further

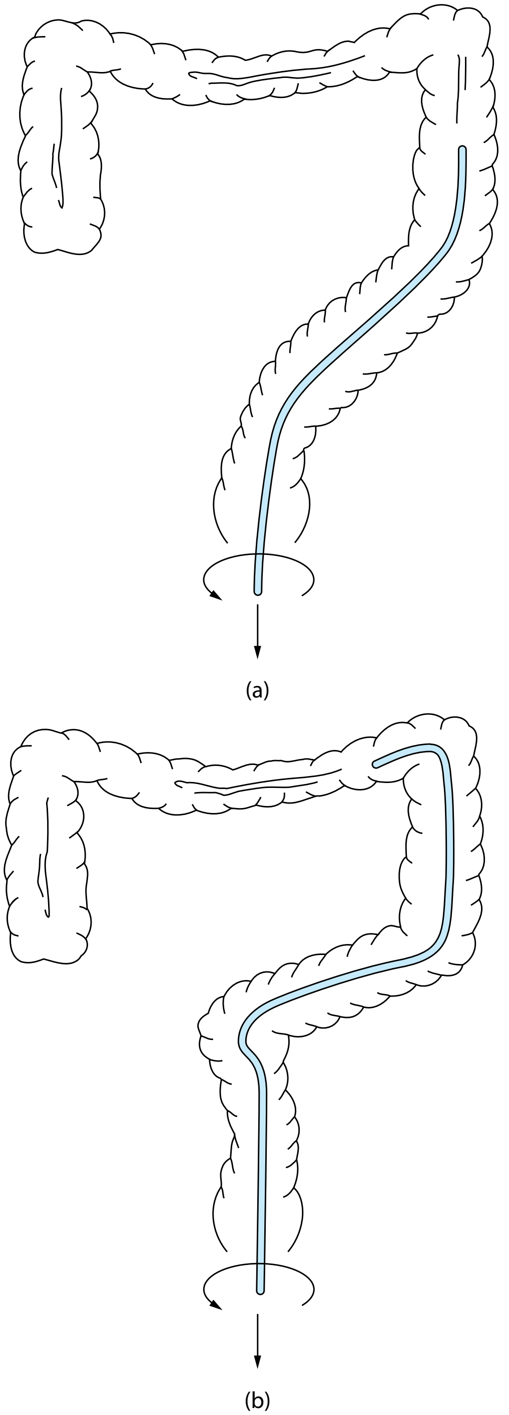


Figure 30.7 Straightening a sigmoid N loop (a) and α loop (b).

loop (Figure 30.8). This again may be overcome by withdrawing the colonoscope and it often helps to turn the patient on to his or her back. Throughout this procedure it is important not to overinsufflate the colon and regular withdrawal of the colonoscope with suction will 'concertina' the colon over the colonoscope. When the hepatic flexure has been reached a useful manoeuvre is to angle the tip of the colonoscope into the ascending colon and apply suction. This will often bring the caecum up to the tip of the colonoscope. If this is unsuccessful then turning the patient on to their right side (facing the endoscopist) may be a useful manoeuvre.

It is very important to recognize when the caecum has been reached and the only two really definite landmarks are the ileocaecal valve and the appendix orifice. Ideally the ileocaecal valve should be entered and a biopsy taken of the ileal mucosa to prove completion of colonoscopy but this is a tricky and time-consuming manoeuvre and few endoscopists would recommend doing this routinely. It must be emphasized that the anal canal and the caecum are the only two parts of the large bowel which can be identified unequivocally during colonoscopy, and even the most experienced colonoscopist will have difficulty in identifying the exact anatomical location of a lesion lying within the colon if it is not in the lower rectum or the caecum.

Complications of colonoscopy

The main complications of colonoscopy are perforation, haemorrhage, bacteraemia and cardiac arrhythmias. Perforation may result at the tip of the instrument or at the apex of a stretched loop (Figure 30.9). Perforation should always be suspected if there is bleeding or if fat can be seen through the wall of the colon. Obviously when small bowel loops can be seen the diagnosis is not in doubt. Perforation can also occur after polypectomy (see section Polyps) but this may be delayed. Likewise hot biopsy (see section Polyps) may lead to delayed perforation. The incidence of perforation is about 0.1% and all patients undergoing colonoscopy should have this risk explained.

Haemorrhage is nearly always a consequence of polypectomy. Significant haemorrhage after an ordinary colonoscopic biopsy is unlikely unless the patient has a coagulation disorder. Bacteraemia has been shown to occur after colonoscopy and it is therefore necessary to use intravenous antibiotics in patients with pacemakers and artificial or diseased cardiac valves. Cardiac arrhythmias commonly occur during colonoscopy but they are usually insignificant and self - limiting. It is however important that patients with cardiac disease should have continuous electrocardiogram (ECG) monitoring and it is essential that resuscitation equipment is present when colonoscopy is being performed.

Colonoscopy with PillCam Colon: capsule colonoscopy

The PillCam Colon (Given Imaging) capsule (31 × 11 mm) introduced in 2008 has two cameras (one at each end), each with a frame rate of four images per second and a battery life of 10 hours. To save power, the device can function in a preprogrammed delay mode keeping the capsule switched off until it reaches the colon. Special software enables the position of the capsule to be monitored during its transit through the gastrointestinal tract to be monitored and helps to identify the optimum time for administration of an additional dose of sodium phosphate to facilitate its transit. Special bowel preparation for capsule colonscopy (CC) is necessary to ensure a clean colon as irrigation for lens cleaning is not possible.

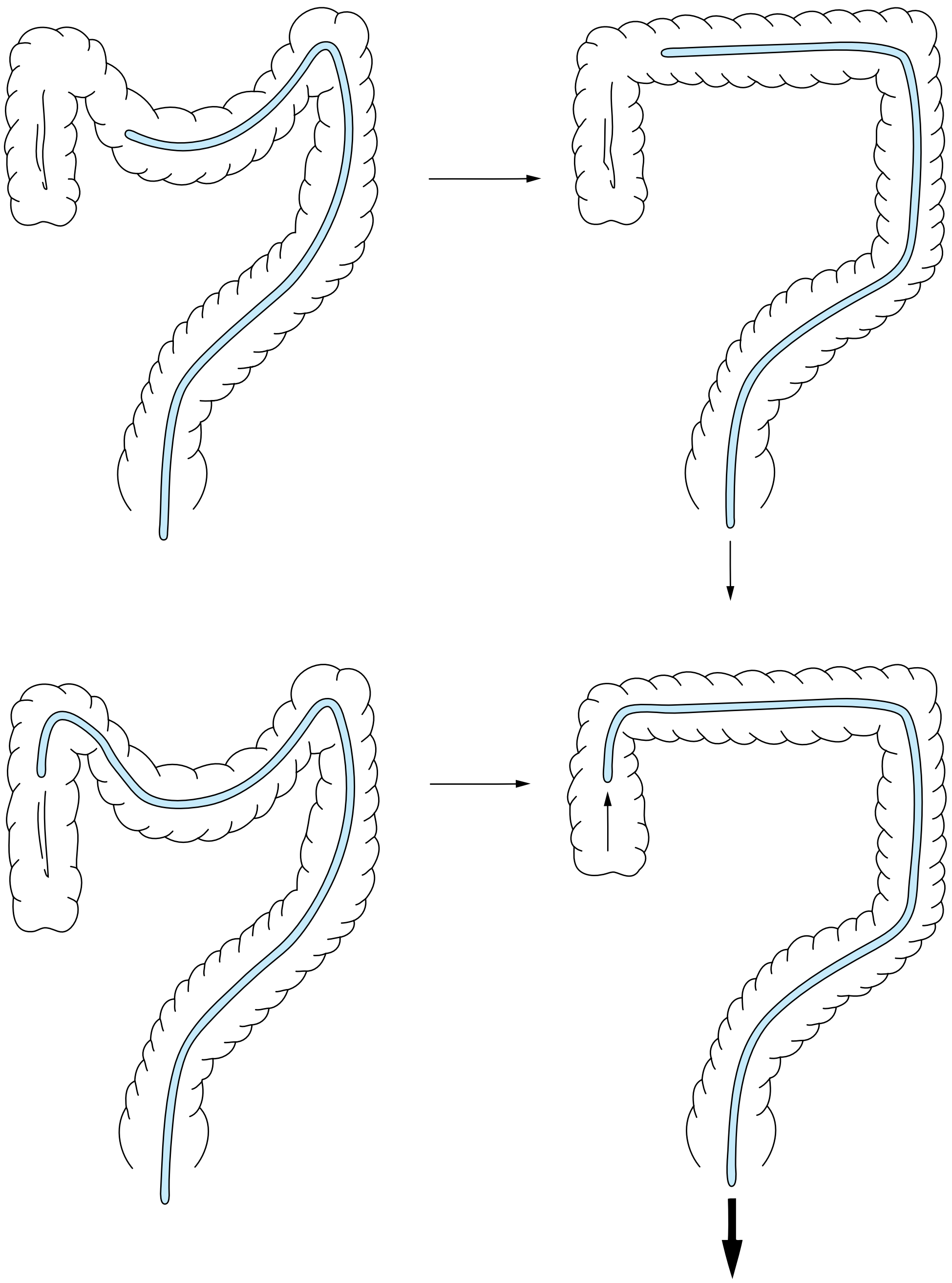


Figure 30.8 Progression of the colonoscope along the transverse colon.

This consists of a combination of diet, PeG-based solution, oral sodium phosphate, oral prokinetic and a rectal suppository. The liquid diet is commenced 1 day prior to the examination and kept until ingestion of the PeG-based solution. The patient ingests the capsule on the day of examination together with a dose of sodium phosphate. A second dose of sodium phosphate may be administered to facilitate capsule transit.

Detection of colonic polyps and colon cancer by CC

In the first multicentre, prospective study of PillCam Colon Ce, which involved 84 patients, CC provided a diagnostic yield of 76% compared with 80% achieved by flexible colonoscopy in the detection of polyps of any size. CC identified major lesions (polyps ≥ 6 mm) in 14 patients, whereas colonoscopy identified major lesions in 16 patients. The sensitivity of the CC reported by the three investigators who examined the data (in blinded fashion) ranged from 50% to 70%, whereas specificity ranged from 83% to 100%. In a single - centre, prospective study of 36 patients, CC detected polyps in 64% of patients compared with 69% detected by colonoscopy. For major lesions, the sensitivity and specificity were 60% and 73%, respectively. Currently, CC is used to complement flexible colonoscopy, being indicated in the rare instances when colonoscopy is refused by the patient or contraindicated because of comorbidities that preclude safe sedation of the patient. However, CC in its present form has major limitations: many patients are unwilling to undergo the procedure

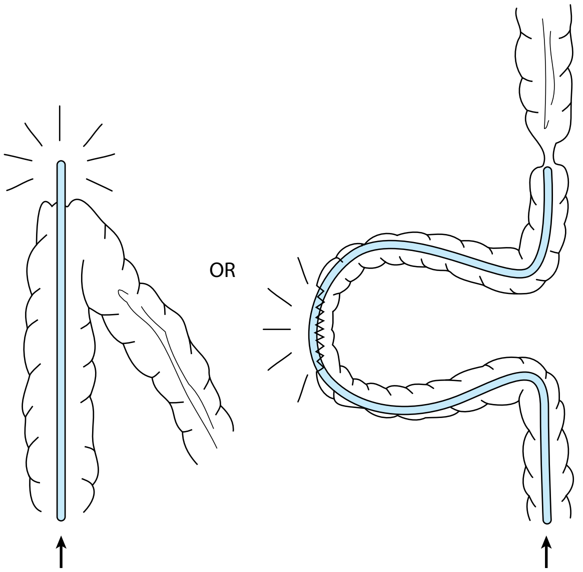


Figure 30.9 Colonoscopic perforation.

because of the complexity of the bowel preparation regimen (as the colon has to be completely clear of all faecal matter), passive transit and inability to biopsy lesions. In addition, technological improvements are necessary to enable the endoscopist to locate lesions and determine the size of polyps. Barring technological developments to overcome these limitations, it seems unlikely that CC will play a significant role or replace conventional colonoscopy in colorectal cancer screening in the short to medium term. The future in this respect lies in microrobotic systems capable of locomotion, imaging and biopsy.

Plain radiology

Plain abdominal radiograph films are particularly useful in the emergency situation. In the patient with mechanical large bowel obstruction, a typical gas pattern will be seen with obvious haustration (Figure 30.10). Pseudo-obstruction may be difficult to distinguish from mechanical obstruction although, in the former, the colonic dilatation usually ends in the region of the proximal descending colon. Obstruction due to volvulus of the sigmoid colon or the caecum will show a typical pattern and in patients with ischaemic colitis typical 'thumb printing' is seen when plain abdominal radiographs are obtained. In the emergency situation an erect chest radiograph should also be carried out. This can be scrutinized for gas under the diaphragm. The plain abdominal radiograph is also essential in patients with an acute exacerbation of ulcerative colitis or Crohn's colitis as this will help in making the diagnosis of toxic megacolon.

Barium enema

The mainstay of radiological investigation of the large bowel is the barium enema. Nowadays, this is always carried out as a double contrast study, i.e. with barium and insufflated air. This coats the colonic mucosa with barium allowing radiologists to detect small lesions and mucosal ulceration. Barium enemas are excellent for diagnosing carcinoma of the colon (Figure 30.11) but caution must be exercised in interpreting appearances in the sigmoid colon and in the caecum. In a redundant sigmoid colon with multiple overlapping loops of bowel, particularly if affected by severe diverticular disease, lesions may be missed (Figure 30.12). Conversely, there tends to be a problem with over-reporting carcinomas in the caecum where prolonged spasm may mimic suspicious appearances (Figure 30.13). Thus in the patient with suspicious symptoms who has a normal looking sigmoid colon, a flexible sigmoidoscopy is mandatory. Likewise in the asymptomatic patient who is not anaemic, a diagnosis of carcinoma of the caecum should be confirmed by colonoscopy unless the appearances are absolutely unequivocal.

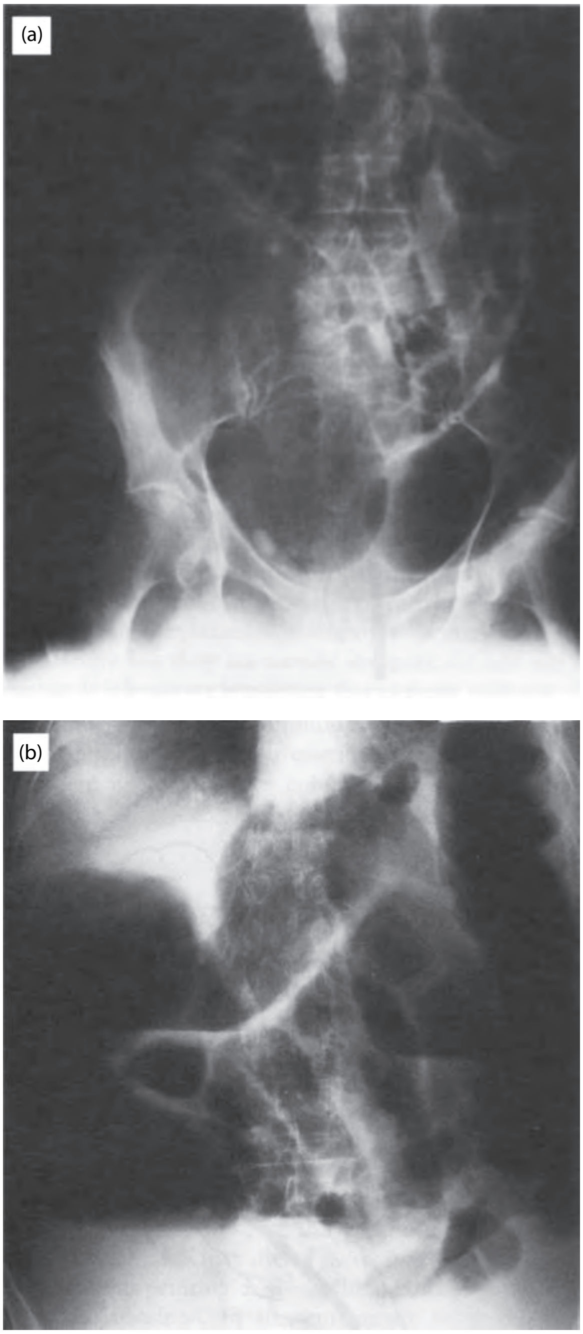


Figure 30.10 Plain abdominal radiographs showing large bowel obstruction: (a) supine film, (b) erect film, showing a fluid level in the caecum.

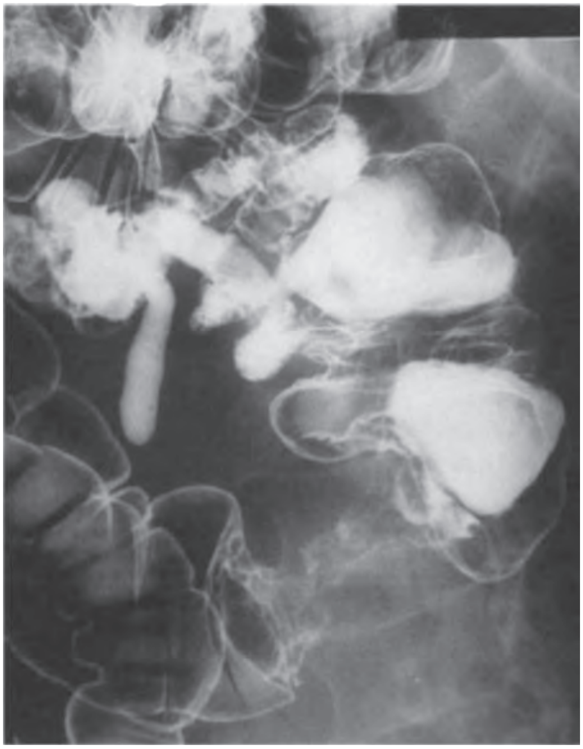


Figure 30.11 Barium enema appearances of a colonic cancer with typical 'apple core' shouldering.

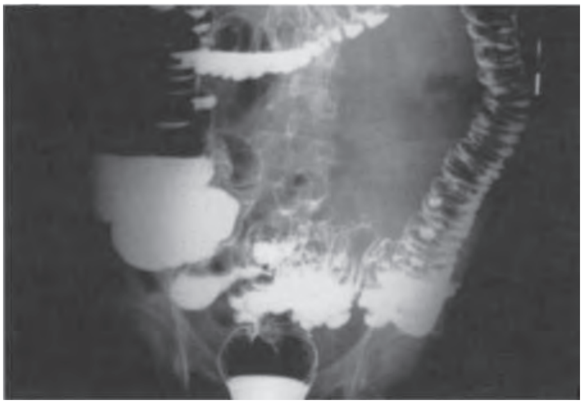


Figure 30.12 Severe sigmoid diverticular disease on barium enema. In this instance a carcinoma was missed.

Strictures in the presence of diverticular disease are particularly difficult to interpret and may require direct endoscopic visualization and biopsy. Small polyps may be missed on barium enema and conversely diverticula or small particles of faecal matter may be misinterpreted as polyps. It should be stressed that both barium enema and colonoscopy are highly operator dependent and both can miss quite significant lesions. Thus in the patient who presents a diagnostic dilemma it may be quite justifiable to use both investigations.

Complications of barium enema

Serious complications are extremely rare. The incidence of perforation is approximately one in 25 000. When it occurs it is usually associated with passage of the rectal tube and it may occur proximal to a colostomy. Very occasionally perforation may result from preparation of obstructed bowel using a stimulant laxative such as Picolax.

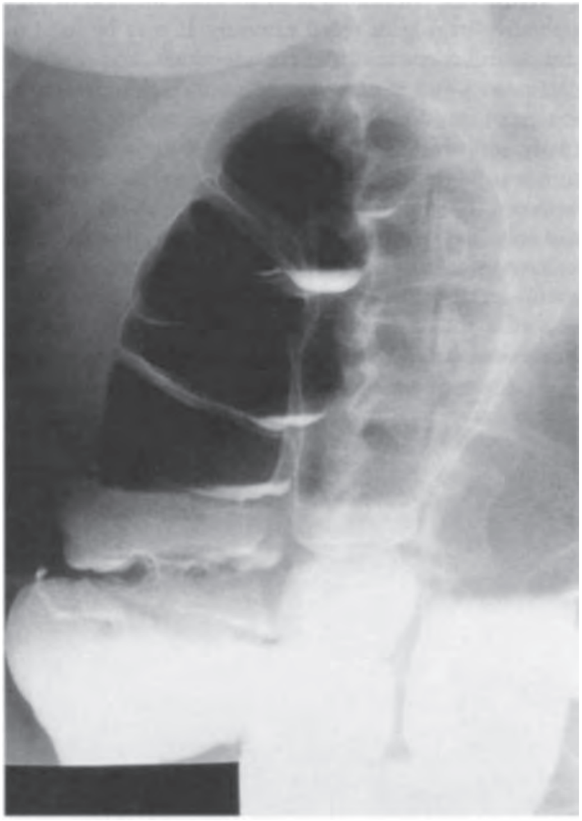


Figure 30.13 Spasm in the caecum interpreted as a stricture on barium enema.

Peroral enema

This technique is sometimes helpful where a standard barium enema has been unsuccessful in demonstrating the caecum. Barium is given by mouth and is monitored by fluoroscopy until it has reached the transverse colon. At this stage air is insufflated per rectum and the right colon is examined. Antispasmodics such as Buscopan may be useful in relaxing the colon for this examination.

Water-soluble enema

If there is a danger of barium entering the peritoneal cavity or if the bowel is obstructed then a water-soluble enema is a useful alternative to a barium. Gastrograffin or Niopam is frequently used. Indications for a water-soluble contrast enema include a suspected anastomotic leak, suspected perforation from or to exclude a pseudo-obstruction. Barium is avoided in the obstructed patient as the study may be followed by immediate surgery which may be compromised by the presence of barium in the colon.

Radio-opaque markers

In order to measure colonic transit radio-opaque markers can be used. Twenty markers are swallowed and in the normal individual the single abdominal radiograph taken on day 5 will show that at least 14 (80%) have been passed. For more detailed information daily radiographs can be taken, or 20 markers of different shapes can be ingested daily on three consecutive days and a single radiograph taken on day 4. In normal individuals the mean total colonic transit time is 30 hours in males and 38 hours in females.

Colonic scintigraphy

Instead of radio-opaque markers it is possible to use material labelled with radioisotopes, e.g. ^{131}I bound to cellulose or indium-^{111} DTPA. The isotope is ingested and the abdomen scanned using a gamma-camera at 6, 24, 48, 72 and 96 hours.

Radiolabelled leucocyte scanning

Intravenous injection of autologous leucocytes labelled with indium-111 can be used to detect active colitis. The leucocytes accumulate in the inflamed mucosa and this relatively non-invasive procedure is useful in assessing the activity and extent of ulcerative colitis and Crohn's colitis. Technetium is an alternative to indium for labelling leucocytes.

Ultrasound

Abdominal ultrasound is widely used for the detection of hepatic metastases both preoperatively and in follow-up for patients with colorectal cancer. It can also be used for the elucidation of abdominal masses and primary colorectal cancers have a characteristic ultrasound appearance. It is not, however, reliable enough to use as a standard investigative technique in a patient suspected to have the disease. Intraoperative contact ultrasound (Figure 30.14) can be used to screen the liver for metastatic disease and indeed this is the most sensitive available method for picking up hepatic deposits (Figure 30.15).

Transrectal ultrasound is an accurate method of preoperative staging of rectal tumours. It may be used to distinguish between benign adenomas and invasive carcinomas and it may also be used to determine the extent of invasion through the bowel wall. It is particularly sensitive in distinguishing between T2 and T3 tumours but not quite so sensitive at distinguishing between T1 and T2 tumours (Figure 30.16). Unfortunately, its sensitivity for picking up lymph node metastases in the perirectal tissues is poor. The most useful rectal probe operates at 7 or 10 MHz and consists of a rotating probe inside a water-filled balloon which can be distended within the rectum to achieve contact between the ultrasound probe and the rectal wall (Figure 30.17).

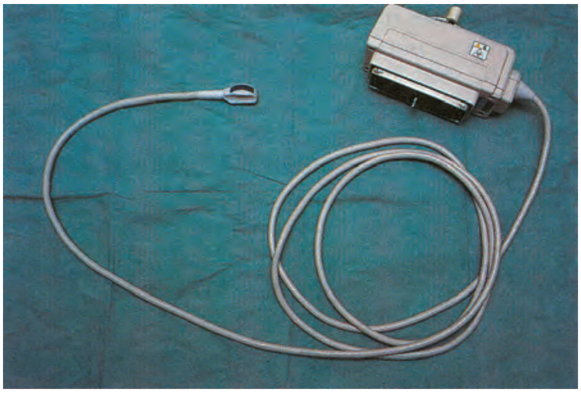


Figure 30.14 Intraoperative ultrasound probe.

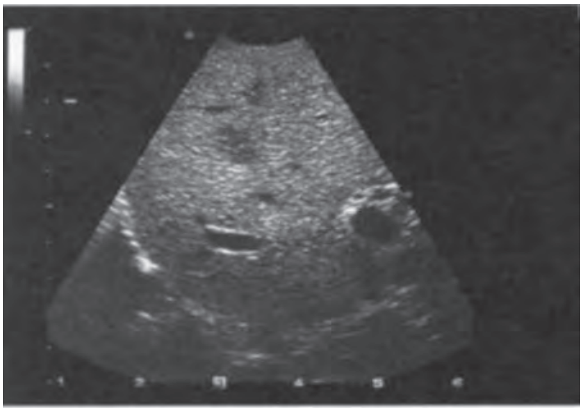


Figure 30.15 Ultrasound appearance of hepatic metastases – image obtained at intraoperative contact ultrasound.

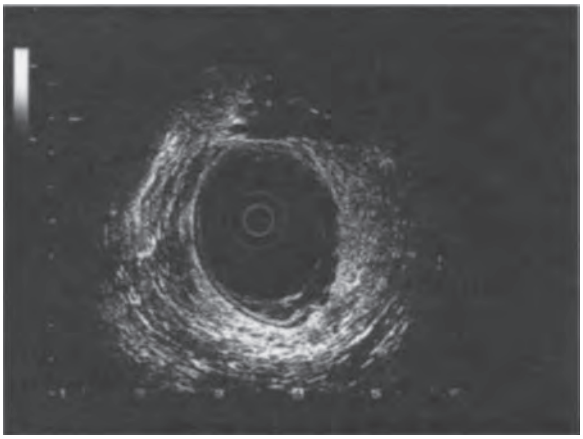


Figure 30.16 Transrectal ultrasound appearance of a rectal tumour. This is a T2 tumour showing breaching of the muscularis propria.

CT scanning

CT scanning is widely used for preoperative staging especially of rectal cancer. It has advantages over ultrasound in that its sensitivity for liver metastases (Figure 30.18) is higher, and an image of the primary tumour in the pelvis can be obtained (Figure 30.19). Again CT can also be used to diagnose abdominal masses and it is of particular value in elderly patients where barium enema and colonoscopy are impractical because of the patients' inability to retain either contrast or air.

A recent development has been virtual colonoscopy where the dataset acquired from a helical CT scan of the abdomen can be reconstructed in order to provide a 'fly through' view of the inside of the colon (Figure 30.20). This procedure shows great promise but like barium enema and colonoscopy it requires full bowel preparation and insufflation of air or CO_2. In addition, because of pooling of liquid in the colon, the CT has to be repeated once with the patient lying prone and again with the patient lying supine. Virtual colonoscopy has been quite widely studied but as yet it has not been shown to be more

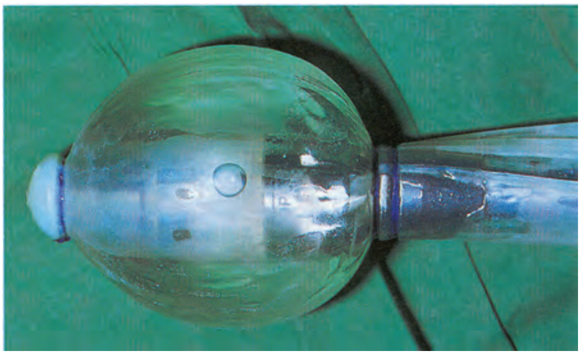


Figure 30.17 Head of a rectal ultrasound probe with the balloon inflated to ensure contact with the rectal wall.

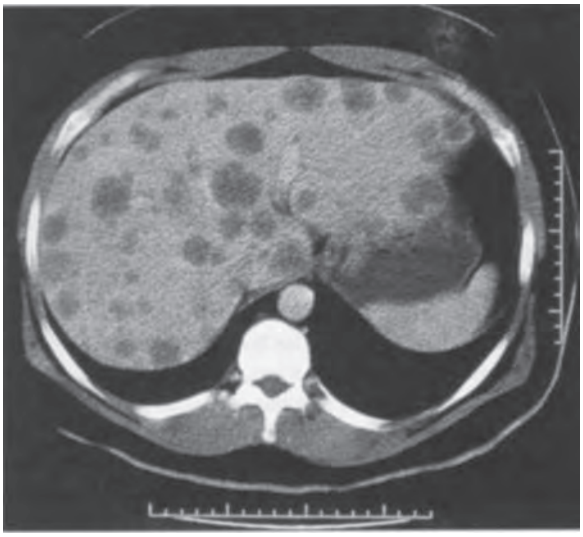


Figure 30.18 CT scan showing multiple hepatic metastases.

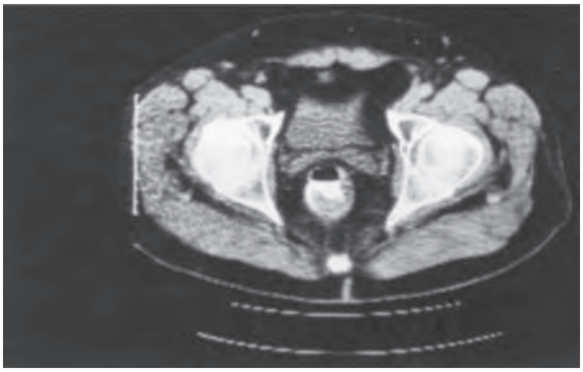


Figure 30.19 CT of rectal cancer showing posterior operable tumour.

sensitive or specific than either a high-quality barium enema or colonoscopy. However, future developments may lead to this technique being able to discriminate malignant lesions from non-malignant lesions with a high degree of accuracy.

MRI scanning

Because of the slow image acquisition time MRI scanning has not supplanted the much more rapid helical CT scanning. However, it does provide particularly good images of the rectum and the latest 1.5T machines with pelvic coils provide very high-definition pictures of the rectum and the mesorectum which has improved preoperative staging considerably (Figure 30.21).

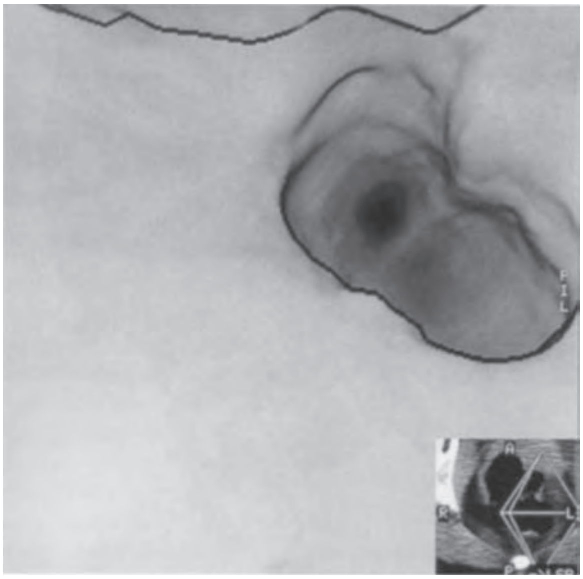


Figure 30.20 CT colography ('virtual colonoscopy').

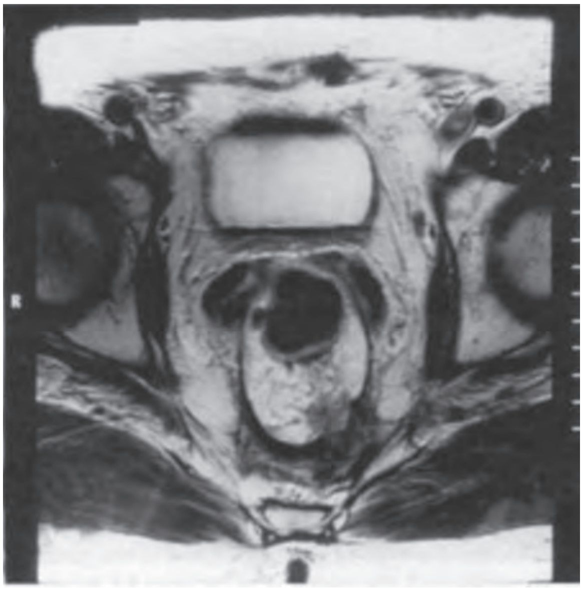


Figure 30.21 MRI of the rectum obtained using phased array pelvic coils.

Functional disorders of the large bowel and pelvic floor

The pelvic floor comprises all the structures that close and support the bottom of the pelvis. These structures provide both dynamic and static support via musculofascial components, enabling it to support, sustain and regulate the canals running through its hiatus. These include the urethra, anal canal and vagina in females, with these organ systems functioning interactively with the supporting pelvic floor.

Gross anatomy

Just inferior to the abdominal cavity and its structures lies the bony pelvic ring, with the levator ani muscles running across this space. The bony borders to this opening traversed by the pelvic floor are the pubic bones anteriorly, ischial spines laterally and the sacrum posteriorly.

In practical terms, the pelvic floor is synonymous with the levator ani because this muscle forms the effective contractile support structure of the region. This muscle takes the form of a broad thin sheet that attaches anteriorly to the posterior surface of the body of the pubis, and is suspended laterally from the pelvic wall to the ischial spines and coccyx, which lie posteriorly. The part of the muscle that attaches to the pubis forms the medial portion of these fibres, and from the pubis this attaches to the perineal body behind the prostate to form the levator prostate. In the female, these medial fibres attach to the lateral vaginal wall to form the pubovaginalis. Other fibres from the pubis pass behind the anorectal flexure, where they fuse with the deep part of the external anal sphincter (EAS) to form the puborectalis. The fibres running from the pubis, together with the fascia covering obturator internus, form the pubococcygeus. The part of the levator ani arising from the lateral wall of the pelvis posterior to the ischial spine is known as the iliococcygeus. Pubococcygeus and iliococcygeus merge imperceptibly with each other as a continuous sheet of muscle. The levator musculature provides support to all the pelvic floor organs and is traversed by the urethra, vagina (in women) and anus.

All the components of the levator collectively contribute to maintaining the position of the pelvic viscera. In females, contraction of the levator ani and its attachments results in the anterior movement of vagina and ESA anteriorly towards the pubic symphysis. On contraction of the pelvic floor, this anterior movement produces occlusive forces on the urethra and rectum, thus contributing to dynamic support of these organs and hence continence. Because the levator muscle complex provides support to all three organ systems (urethra, vagina and anus), its weakness will result in impaired function of any or all of the structures being supported. Muscular dysfunction can result from stretch or tear injuries to the pelvic floor muscles; however, the most common aetiology for muscular dysfunction is a denervation injury from childbirth. Injury to the pudendal nerve can result in dysfunction of the urethral sphincter, anal sphincter, and motor or sensory dysfunction of the perineum. Significant damage to pudendal innervation can result in multisystem dysfunction, causing prolapse and combined mixed incontinence.

Numerous epidemiological studies link pelvic floor disorders with parity. Damage to nerve, muscle and connective tissue may occur after vaginal delivery, all of which may contribute to pelvic floor disorders. With increasing age, latent injuries from vaginal deliveries may manifest as continence disorders, difficulty with urination and bowel motions, and pelvic organ prolapse. Surgery may be a potential curative option, provided the right diagnostic strategies are initially followed. Care must be taken as pathology is often multifactorial, and restoration of anatomy may not always provide positive short- and long-term functional outcomes.

Epidemiology of pelvic floor-related bowel dysfunction

These disorders include several different clinical problems such as faecal incontinence, constipation and prolapse of the rectum. They are often complex and can involve the functions of smooth and skeletal muscles, their nerves and connective tissues. Aetiology can be post-traumatic (anorectal surgery, difficult vaginal deliveries), or from chronic evacuation difficulties. Many cases can be idiopathic.

Epidemiology of faecal incontinence

The true prevalence of the problem varies in every country. Faecal incontinence is frequently underestimated because of the embarrassment and reluctance of patients to discuss the condition, and varying levels of coping. The incidence of faecal incontinence ranges between 0.5% to 5% in most studies. It is higher in institutionalized individuals – around 30% in elderly people and 50% in patients with neuropsychiatric disorders. Gender also plays a significant role. Over the age of 45, women are nine times more likely to suffer with faecal incontinence than men of the same age. This increased risk may be attributable to obstetric injury, which is likely to be the most important aetiological factor in women. Other risk factors for faecal incontinence include diabetes, irritable bowel syndrome (IBS) and anorectal surgery.

Despite traumatic childbirths, which may lead to unidentified sphincter defects, only a minority of women ever become symptomatic early on. These patients may however become symptomatic later in life or with subsequent vaginal deliveries. A varying number of these women will also have sustained pudendal nerve injuries, which may manifest in different ways. The incidence of sphincter injury is significantly higher in patients with perineal tears. Menopause and ageing is also noted to be a risk factor for the development of altered continence. The prevalence of faecal incontinence in women 75 years and older is between 9% and 17% – significantly higher than the general population.

Epidemiology of evacuatory disorders

Constipation is one of the most common complaints of patients seeing their GPs or colorectal surgeon. The prevalence of chronic constipation can range from 5% to 35%. This can be caused by several aetiologies. Obstruction of the pelvic outlet, often known as obstructed defecation syndrome (ODS), is a common cause of constipation and is attributable to muscular dysfunction of the pelvic floor. In a population study performed by Talley *et al*, the prevalence of this condition was 16.5% in women and 5.2% in men. There was a significant increase in prevalence of this form of evacuatory disorder with age.

Multiple risk factors are associated with ODS. Constipation in youth to middle age is approximately three times greater in women, and the prevalence increases in men and women with age. Laxative and enema abuse have also been associated with ODS.

Rectal prolapse

Rectal prolapse occurs when the full thickness of the rectal wall protrudes through the anus. It is the most common type of distal gastrointestinal tract prolapse. If only the mucosa of the rectal wall protrudes through the anus, this is termed a mucosal prolapse. By contrast, full or partial thickness prolapse can also occur in the rectum but not pass beyond the anus. This is termed internal intussusception. The true incidence of rectal prolapse is not known because of under-reporting. It is associated with longstanding constipation, chronic straining, pregnancy, anorectal surgery, female gender, ageing, neurological disease (such as dementia) and mental illness.

Diagnostic modalities in pelvic floor disease

Pelvic floor disease comprises numerous clinical manifestations, depending on the precise nature and combination of pathology of the local anatomy and function. In order to accurately diagnose these disorders, a number of radiological techniques are utilized to assess structure and function, statically and dynamically. In assessing patients, there is general recognition that the majority of them have a multiplicity of pelvic floor and perineal soft-tissue abnormalities across compartments. Dynamic imaging modalities are therefore required to define the real-time integration of these anomalies and to highlight their significance in each case, particularly where there is clinical or radiographic evidence of a dominant pathology and where corrective surgery is contemplated.

Ultrasound scanning, X-ray dynamic contrast radiography and MRI form the mainstay of anatomic and functional assessment. Transperineal ultrasound can be used to assess the component parts of the anterior, middle and posterior pelvic compartments to define their interaction during straining and simulated defecation in patients with symptoms of evacuatory difficulty.

Pelvic ultrasound

Ultrasonography of the pelvic floor has assumed a central role in the diagnostic work-up of pelvic prolapse syndromes, obstructive defecation and faecal incontinence. It can be a useful first-line imaging approach as it helps discriminate patients with anatomical abnormalities who are likely to benefit from surgery and those requiring conservative management. More recently, dynamic anorectal ultrasonography and MRI have been used successfully in evaluating these disorders. Introital ultrasound and, more importantly, endoanal ultrasound have become accepted approaches for assessing anal sphincter pathology.

Pelvic MRI

MRI has an important role in analysing the pelvic floor musculature. In the supine position, the pelvic floor is dome shaped at rest. During voluntary contracture, the levator straightens out and becomes more horizontal. While bearing down, the muscles descend a variable amount and become basin shaped, with widening of the genital hiatus. MRI has previously shown inward movement of the levator on contraction and outward movement with straining. MRI scanning of the levator ani can assist in the recognition of asymmetry between the two sides, demonstration of the relationship with the obturator internus muscle, visualization of defects and, if the appropriate images are taken, determination of muscle volume.

Dynamic contrast proctography

Dynamic contrast proctography (henceforth referred to as defecating proctography) has traditionally been used for investigating intraluminal anorectal abnormalities, with well - established criteria for diagnosing rectal prolapse and intussusception. In its classic description, mucous prolapse is defined as a rectal wall infolding <3 mm thick confined to the anterior or posterior margin, which does not show a tendency to migrate distally on straining. Conversely, a circumferential infolding of the rectal wall <3 mm that descends toward the anal canal is defined as intussusception. This is described as intrarectal when it remains within the rectum and intra-anal if its apex penetrates the anal canal. Defecating proctography can determine morphological features such as thickness, depth of descent and the point of inversion from the anal verge.

In addition, intussusception can be graded by using a three-point scale, depending on the involved rectal wall appearance at the end of evacuation:

- grade 1 intussusception is a 3–5 mm thick intraluminal filling defect that has a mucosal component only
- grade 2 is 5–10 mm thick and includes both mucosal and mural components
- grade 3 or full-thickness intussusception is assumed if the prolapsed folds are thicker than 10 mm, penetrate the anal canal and appear to impede the expulsion of rectal content.

A complete external rectal prolapse is diagnosed when the entire thickness of the rectal wall is extruded through the anal canal. This can be reducible or irreducible.

A rectocele appears as any rectal protrusion anterior to a line extending upward through the anal canal (Figure 30.22). Its depth, either at rest or on straining, should be measured as the shortest distance from the deepest margin of where it actually reaches the anterior rectal wall and the expected line mentioned above. Rectoceles measuring <2 cm deep are graded as small, moderate if measuring 2–4 cm, and large if >4 cm. With the introduction of open-configuration MRI scanners, image acquisition in the vertical or even sitting position is now possible and provides a global view and analysis of the musculature, pelvic viscera, anorectal angle, pelvic floor muscle function and pelvic floor descent. With this technique the rectal wall is seen even more clearly, making intussusception and rectocele evaluation even more reliable. Dynamic MRI has been found to be at least equivalent to conventional defecography in all aspects tested. This technique has the advantage of showing sphincter function and global motility of the pelvic floor while avoiding radiation exposure, constituting a significant aid to surgical and therapeutic planning.

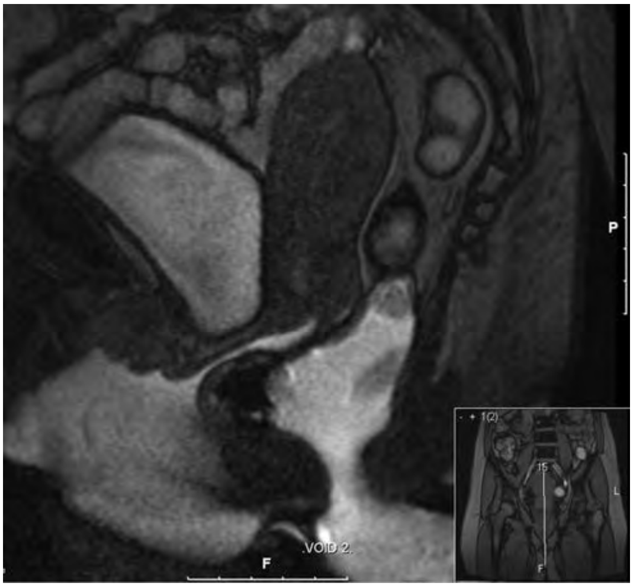


Figure 30.22 Magnetic resonance defecating proctogram showing pelvic floor descent and rectocele.

Endoanal ultrasound vs endoanal MRI

Anal incontinence is a complex and debilitating condition producing considerable morbidity in those affected. Faecal continence depends on intact function of the anorectal sphincter, appropriate consistency of faeces and maintenance of higher mental function.

Anorectal sphincter function may be impaired by disruption of the EAS or internal anal sphincter (IAS), atrophy of the anorectal sphincter mechanism or a combination of all three. Disruption of the EAS occurring during vaginal delivery is the most common traumatic cause of anal incontinence. IAS disruption is less common after local trauma, most commonly as a result of surgery. Isolated disruption of the IAS is not seen after childbirth but often occurs combined with external sphincter damage. Sphincter atrophy is generally related to pudendal neuropathy often presenting as delayed-onset incontinence. Along with functional assessment, anatomical assessment of the anorectal sphincter as well as accurate description of the site and cause of sphincter disruption are fundamental to patient management.

Endoanal ultrasound and endoanal MRI are the principal imaging modalities employed to make an anatomical assessment of the anorectal sphincter. Studies in healthy volunteers have demonstrated the relative merits of each technique. Endoanal ultrasound performed with the patient prone or in the left lateral position using a 7 MHz or 10 MHz rotating transducer allows 360° visualization of the anal canal. Endoanal ultrasound is widely available and provides excellent visualization of the IAS enabling the focal defects and thinning to be detected. Incomplete visualization of the external sphincter, difficulty in distinguishing pathological defects from the normal asymmetry, reliance on operator dependence and poor interobserver variation have been cited as limitations of the technique. Images of the anal sphincter complex obtained using endoanal MRI are thought to be superior to MRI performed with a body coil because of increased signal-to-noise ratio, resulting in high spatial resolution images. However, the limited availability of endoanal coils outside specialist units has resulted in less widespread familiarity with this technique. Endoanal MRI allows the comprehensive assessment of atrophy and focal defects of the external canal; however, the internal sphincter is less well defined. Several studies comparing the diagnostic accuracy of endoanal ultrasound and endoanal MRI have been performed. The cohorts are generally small. The use of surgery as the diagnostic 'gold standard' results in the over - representation of patients with disruption of the EAS, as these alone are candidates for surgical repair. Local practice and expertise is often reflected in the findings of these single - centre studies.

Magnetic resonance defecography

Defecography is one of the most important examinations to be performed in patients with pelvic floor dysfunction. It can provide dynamic and static images of pelvic floor structures during rest, squeeze and evacuation. The examination depends on voluntary control of the pelvic floor and passive emptying of the rectum.

Defecography is the dynamic study of evacuation. Wallden was the first to investigate the relationship between changes in pelvic dynamics, rectocele, enterocele and symptoms of obstructed defecation. Defecography allows a closer physiological evaluation of the evacuation process than other diagnostic tools. It also provides measurements in the resting, squeeze and bearing-down positions to show both anatomic and physiological abnormalities. Defecography may be indicated in patients with chronic constipation to evaluate obstructed defecation, which can be the result of rectocele, sigmoidocele, enterocele, internal mucosal intussusception, anismus and increased perineal descent. Treatment should still be planned in combination with patient symptoms. Performing this study by MRI is an accurate method for evaluating morphology and function of anorectal and pelvic muscles and organs. The dynamics of the pelvic floor may be evaluated in multiple compartments in high-resolution images and video (dynamic) mode.

Unlike earlier forms of imaging such as conventional X-ray cinedefecography, MRI makes it possible to view the pelvic floor in its entirety and in multiple compartments both at rest and dynamically, thus providing information essential for surgical planning and choice of treatment approach. With MRI, the opening of the anal canal and the anorectal angle can be evaluated during sphincter contraction and evacuation, and the elimination of endorectal contrast quantified. The rectal walls can also be clearly observed, and disorders such as intussusception and rectocele may be identified. In addition, it is possible to view the descent of anterior, mid- and posterior compartments of the pelvis. With the addition of an endorectal or endoanal coil, the structure of the anal sphincters may also be assessed. MRI can be performed in the upright or sitting position utilizing an open-configuration system. This has the advantage of simulating the physiological position of defecation, while sacrificing spatial resolution due to necessary attenuation of the MR signal with the open magnet.

Defecography with MRI provides accurate evaluation of the morphology and function of the anorectal and pelvic muscles and organs involved in pelvic floor dynamics. The spatial resolution is high enough to view the relevant morphological structures and dynamics of the pelvic floor, demonstrating relevant pathologies affecting the defecatory mechanism.

Anorectal manometry

Functional anorectal manometry (ARM) testing was developed during the early 1960s. ARM allows a better understanding of the physiology of continence and defecation, with a clearly defined role in the diagnostic work-up of diseases of the pelvic floor.

The anal canal is a narrow passage measuring 2–5 cm long, which is normally closed by the IAS and EAS. The IAS is smooth muscle, maintaining approximately 80% of resting anal tone, whereas the EAS, a striated muscle, is responsible for the remainder. The puborectalis muscle, a U-shaped component of the levator ani complex, blends with the upper end of the EAS and maintains an acute rectoanal angle at rest. Rectal distension evokes involuntary relaxation of the IAS and increases rectal contractility, thereby inducing the desire to defecate. If convenient, defecation is then initiated by relaxation of the anal sphincters and pelvic floor muscles. Conversely if inconvenient, the rectum can relax to accommodate stool, with the sense of urgency subsiding. The EAS and pelvic floor muscles, being striated muscles, can be contracted voluntarily to preserve faecal continence. ARM requires a catheter (which comes into direct contact with the patient), pressure transducers, a recording system to store the digitized signals and software to analyse the recordings.

Multichannel water-perfused catheters used to be the most common tools for performing anal manometry. These have lately been replaced by solid-state or air-charged catheters, which are considerably easier to manage and maintain. Measurements can be performed as a continuous pull-through, or at set levels within the anal canal (station pull-through). The continuous pull-through technique requires the catheter to be withdrawn at a continuous speed from the anal canal, which can provide a detailed recording of pressure profiles, and computer-based three-dimensional representations of the anal canal. With the widespread availability and ease of use of endoanal ultrasound, station pull-through has become more common, with anal canal pressure measurements performed at 1 cm increments in the anal canal. This method provides a more accurate assessment of anal pressures since there is a stabilization period between each reading, thereby reducing artefact. This is the method that the authors perform, with the catheter tip placed 5 cm cephalad from the anal verge. Anal pressure measurements are then taken at 1 cm intervals after a stabilization period of 10–20 seconds. Rest and squeeze pressure are recorded at each centimetre up to 1 cm from the anal verge. Anal pressures, capacity, sensation and rectoanal inhibitory reflex are tested. The rectoanal inhibitory reflex is thought to be central to the 'sampling effect' of the rectum. Rectal distension with small volumes should cause a contraction of the EAS followed by a pronounced IAS relaxation. This reflex is absent or diminished in patients with Hirschsprung disease.

The mean resting pressure in healthy volunteers ranges from 40 to 70 mmHg. The IAS generates around 60% of resting pressure, with the external sphincter and surrounding soft tissue contributing the remaining 40%. Maximal squeeze pressure in healthy individuals should be between two and three times baseline resting value. The EAS is the main contributor to these pressures. Anal resting and squeeze pressures are frequently reduced in faecal incontinence. Among patients with weak or normal anal pressures, other factors such as diarrhoea and disturbances of rectal compliance/sensation may also contribute to faecal incontinence. ARM can also be useful to discriminate defecatory disorders from other causes of chronic constipation. The rectal balloon expulsion test for example is a useful screening test for a functional defecation disorder. Features of evacuation disorders on ARM include impaired relaxation or paradoxical contraction of the anal sphincter (i.e. anismus) or inadequate increase of rectal pressure during simulated evacuation. Anal resting tone may be increased in patients with a functional defecation disorder, as well as in patients with anal fissures.

Colonic transit studies

These are commonly performed with a radioactive marker or radio-opaque capsules (Figure 30.23), commonly known as a Sitzmark test. This test is often useful in patients who are otherwise well, but suffer with chronic constipation refractory to medical management. The test allows a clinician to differentiate between normal transit and slow transit constipation. In this procedure, the Sitzmark capsule is ingested, and abdominal radiography performed 5 days following its administration (Figure 30.24). In normal controls, most markers are eliminated by day 5. If more than 20% of the markers remain in an equally scattered fashion, then slow transit constipation exists. If the markers are confined to a particular area of the colon, this may imply some form of obstruction at that point, which needs further investigation. If no obstruction is found, more detailed radioactive marker - based assessment with multiple abdominal radiographs over

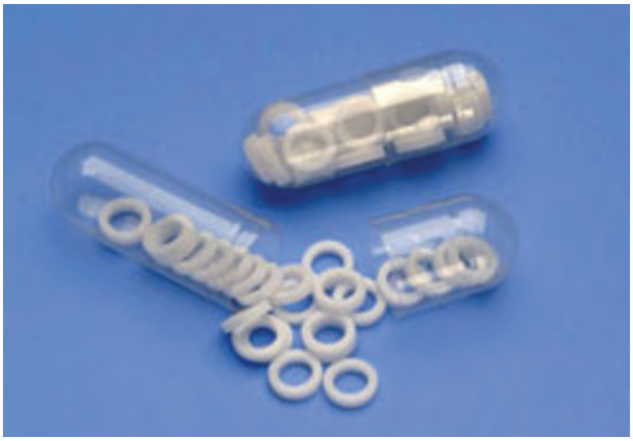


Figure 30.23 Sitzmark capsules.

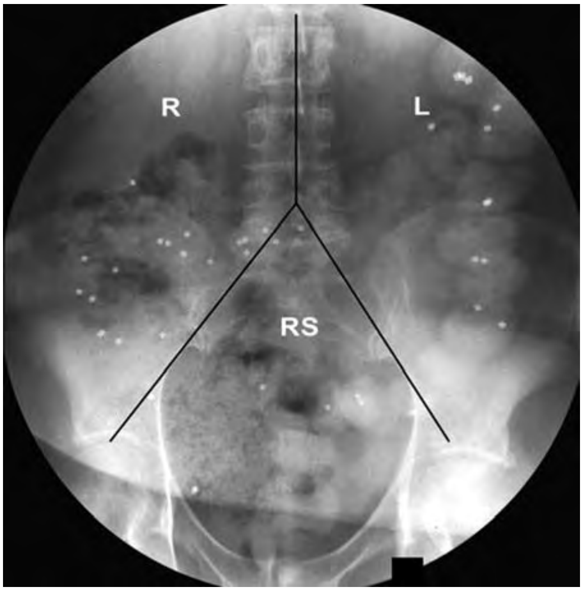


Figure 30.24 Sitzmark's test with markers in the right hemicolon (R), left hemicolon (L) and rectosigmoid (RS).

time may need to be performed, in order to better define the affected segment of colon.

Functional manifestations of pelvic floor disease

Pelvic floor disease can manifest in a number of ways. This can comprise both functional and anatomical pathology, resulting in disorders of urinary function, faecal function and pelvic organ position. Varying studies have been performed to provide a better understanding of the coexistence of urinary, faecal and genital symptoms. There is a high incidence of coexistence of incontinence and genital prolapse. In patients seen with faecal incontinence, 20–55% also complained of urinary incontinence, with 7–20% also complaining of genital prolapse. Conversely, of patients presenting with rectal prolapse, over 60% also complained of urinary disturbance and 30% of genital prolapse. The frequent coexistence of urinary and colorectal dysfunction indicates the need for a multidisciplinary approach to the evaluation of patients with these conditions.

Urinary incontinence

Urinary incontinence is the involuntary loss of urine. Stress urinary incontinence is the involuntary loss of urine that occurs during physical activity such as coughing, sneezing, laughing or exercise. It is a distressing problem of bladder storage in which the strength of the urethral sphincter is diminished and unable to prevent urine flow when there is increased abdominal pressure. This is often, but not exclusively, the result of weakness of pelvic floor support with resulting rotational descent of the bladder neck and proximal urethra during increases in intra-abdominal pressure. Urge incontinence is the sudden involuntary contraction of the bladder causing an immediate unstoppable urge to urinate. Urinary incontinence can be of the stress, urge or mixed variety. Epidemiological studies have estimated prevalence rates of between 10% and 40%. The EPICONT study in Norway, which involved over 20 000 women, noted a gradual increase in prevalence until age 50, when it seems to plateau at about 30%, rising again at age 70.

Faecal incontinence

Faecal incontinence is the involuntary or inappropriate passage of flatus, liquid or solid stool that is a crucial and hygienic problem. Similarly to urinary incontinence, this can be stress, urge or mixed incontinence. In addition, incontinence can also be passive, occurring without one's knowledge. The pathological basis of faecal incontinence can be muscular, neurogenic or idiopathic. Patients may have a decrease in sensory acuity in the anal canal at the anal margins. Often the patient cannot distinguish between the presence of flatus or faeces in the anorectum. This prevents the normal sampling reflex from working appropriately, resulting in, initially, involuntary loss of flatus and liquid stool, progressing to incontinence of formed stool. Faecal incontinence is often underestimated because of the embarrassment and reluctance of patients to discuss this condition. The social stigma and embarrassment of being incontinent of faeces can lead people to severely limit their activity. Incontinence of stool or flatus occurs at least weekly in 2% of adults. This rate may be higher in women, especially in those with obstetric-related structural sphincter damage. With increasing age, prevalence of faecal incontinence also rises, which may reflect degenerative changes of the sphincter mechanism over time. The consequences of faecal incontinence on a patient's quality of life may be substantial. Loss of self-esteem, difficulty travelling or maintaining employment, and strain on personal relationships may adversely affect patients and their families, with economic consequences borne by individual and state.

Pathophysiology of faecal incontinence

Normal defecation and continence require a complex interplay of several important factors. The rectum, which functions as the organ of storage, requires appropriate distensibility, compliance and sensory ability. The passive musculature which maintains continence should be of sufficient calibre and contractility, providing an adequate resting pressure to hold back stool in the rectum. The voluntary muscles should also be of normal length and thickness and contract appropriately, maintaining an acute anorectal angle, so that the faecal bolus is supported over the levator muscle and not directly over the anal canal. The outflow system should be intact, and stool transit time should also be within normal range.

Denervation has been found in the pelvic floor muscles of patients with constipation and defecation disorders that cause excessive straining. These patients often show the physical sign of perineal descent. The terminal portion of the pudendal nerve in an adult is approximately 9 cm. In patients with abnormal descent of 2–3 cm, a stretching force to the distal nerve of 20–30% is therefore exerted. Stretching of these nerves during straining may lead to secondary neuropathic damage with consequent pudendal neuropathy.

Patients complaining of faecal incontinence typically present in one of two ways. They may have a significant obstetric history with prolonged labour, instrumental delivery and sphincter injury. Injuries may have been occult and never treated, or surgically repaired either at presentation or in a delayed fashion. The other presentation is a patient with years of forceful straining at defecation. Patients may be aware of the urge to defecate, but are unable to defer or prevent the passage of faeces. On the other hand, patients may complain of leakage in a passive fashion, with faecal incontinence occurring while asleep or unawares.

Management of faecal incontinence

Management of faecal incontinence is primarily by conservative means, including dietary and lifestyle changes, antidiarrhoeal medication, biofeedback therapy, absorbent pads and anal plugs. Surgical options include sphincteroplasty, postanal repair and, more recently, sacral nerve stimulation (SNS). SNS is a minimally invasive technique that allows modulation of the nerves, and therefore muscles, of the pelvic floor. It is achieved via the application of an electrical current to a sacral nerve by insertion of an electrode through the corresponding sacral foramen (Figure 30.25). Sacral nerve stimulators were first implanted in 1981 for the treatment of urinary urge incontinence, with SNS first used for the treatment of faecal incontinence in 1995. In its current form, SNS involves extradural stimulation within the sacral canal, which has reduced complications compared with previous transcutaneous, transvaginal and transrectal techniques. SNS involves a testing phase known as peripheral nerve evaluation (PNE). PNE determines the feasibility of electrode implantation into the sacral foramina, followed by a 2–3 week period of stimulation with a temporary electrode to assess the potential benefits of SNS. This allows identification of patients who are likely to have a positive response to SNS, into whom a permanent electrode can be inserted.

Recent studies and meta-analyses (Table 30.1) have shown SNS to be a safe and highly effective treatment for faecal incontinence, improving both functional and quality of life outcomes in patients where maximal conservative therapy has failed. There is generally an increase in both resting and squeeze pressures following successful SNS implantation, although the significance of this is unclear. The resultant improvement in continence may be from an augmentation action on the sphincter complex and puborectalis, as well as modification of afferent and efferent neurological pathways in bladder and bowel function. SNS influences rectal sensory function, with patients demonstrating increased rectal sensitivity to balloon distension testing. This improved sensation may contribute to an increased awareness of rectal content and hence continence. SNS has also been shown to be effective in unifocal as well as multifocal sphincteric defects.

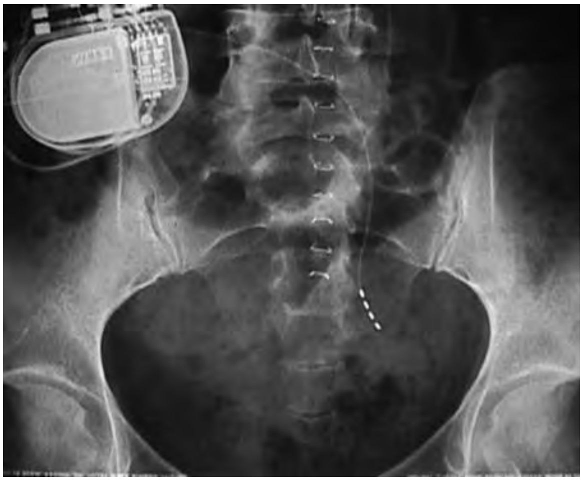


Figure 30.25 Permanent sacral nerve implantation with a tined lead sitting in the left S3 foramen. and an implantable pulse generator.

Table 30.1 Results of a meta-analysis of sacral nerve stimulation (SNS) vs maximal conservative therapy (MCT)

Outcome of interest	No. of studies	No. of patients		WMD
		MCT	SNS	
Functional outcomes				
Incontinence episodes/week	28	622	574	-6.83
Incontinence scores	14	289	272	-10.57
Deferring defecation (minutes)	9	165	158	7.99
SF-36 outcomes				
Physical functioning	7	102	87	11.99
Social functioning	7	102	87	20.91
Role physical	7	102	87	33.82
Role emotional	7	98	87	18.48
Mental health	7	102	87	13.43
Vitality	7	102	87	10.77
Bodily pain	7	102	87	7.99
General health	7	102	87	14.92
FIQL outcomes				
Lifestyle	9	199	169	1.23
Coping/behaviour	9	199	169	1.28
Depression/self-perception	9	199	169	1.16
Embarrassment	9	199	168	1.41
Anal manometry				
Resting pressure (mmHg)	28	613	440	6.40
Squeeze pressure (mmHg)	29	632	455	16.19
Rectal sensitivity				
Threshold volume (mL)	22	462	391	-6.53
Urge volume (mL)	21	441	370	-7.22
Max tolerable volume (mL)	20	406	334	-5.33

FIQL, faecal incontinence quality of life; WMD, weighted mean difference.

As well as its benefits vs conservative treatments, SNS has been demonstrated to be cost effective and less morbid relative to other surgical interventions, including the artificial bowel sphincter, dynamic graciloplasty and end stoma.

Disorders of defecation

This group of disorders is often described as constipation by patients. Constipation, with its associated symptoms, is a common bowel problem. Constipation itself is a misleading name as it does not necessarily mean the passage of hard or infrequent stools. It can also mean excess defecatory strain, or passage of stools that are too small/hard, described as 'rabbit pellets'. The most common symptoms associated with constipation are a feeling of having to strain at stool, abdominal pain or bloating, and dissatisfaction or a feeling of incomplete rectal emptying after passing stool.

The causes of constipation can be varied and multifactorial, being a symptom of organic disease of the gastrointestinal tract. Patients with constipation can be divided into two main groups – those with intestinal dysmotility/impaired transit and those with outlet obstruction. Broadly, the causes can be categorized as (1) systemic factors that affect normal colonic and rectal function, (2) primary colonic and anorectal factors and (3) psychological factors.

ODS is a common entity among the disorders of defecation. It is under-reported, estimated to affect between 15% and 20% of the adult female population. This subset of patients may respond well to surgical management. Known anatomic findings associated with ODS include rectocele and rectal intussusception. It is this redundancy of the distal rectum that has been shown to be amenable to surgical resection with restoration of normal rectal anatomy and resolution of symptoms. Patients with ODS may also suffer with anismus, otherwise known as paradoxical puborectalis contraction, which manifests as a failure of the puborectalis muscle to relax and open the rectal angle necessary for smooth defecation.

Hypothyroidism, diabetes, hypercalcaemia, connective tissue disorders such as scleroderma and amyloidosis can all contribute to constipation. Neurological disease, congenital or acquired, may also be associated. In addition, many drugs cause constipation as a side effect, and most drugs aggravate constipation in some patients. These may include opiate-based analgesics, anticholinergics, antidepressants, beta-blockers, calcium antagonists and diuretics.

Surgical management of obstructed defecation syndrome

Stapled transanal resection of the rectum (STARR) has been proposed by Longo for the treatment of ODS. It involves a double-stapling technique performed transanally to produce a transanal full-thickness rectal resection with the goal of correcting structural/anatomical abnormalities associated with ODS, leading to restoration of rectal flow, normal rectal wall thickness and compliance, correction of rectocele, and correction of rectal intussusceptions. It has been performed for over 6 years now, and numerous publications have demonstrated the safety and efficacy of the procedure in the short–medium term. Rectoceles themselves can of course be treated transvaginally in women to good effect, but there are concerns about long-term efficacy, dyspareunia and wound problems. The European STARR registry, established in 2008, has tracked over 3000 patients who have had the procedure, with good effects maintained at 12 months of follow-up. Complications were reported in 36%, with the majority complaining of urgency, followed by bleeding, septic events, staple line complications and incontinence.

Rectal prolapse

The definition of rectal prolapse is the protrusion of the full thickness of the rectal wall through the anus (Figure 30.26). This protrusion differs from mucosal prolapse and internal intussusception. In cases of mucosal prolapse, only the inner mucosal layer protrudes through the anus, leaving the muscular layers behind. In rectoanal intussusception the prolapsed tissue may be partial or full thickness, but remains confined to the rectal lumen. The type of prolapse will help direct the appropriate therapeutic option.

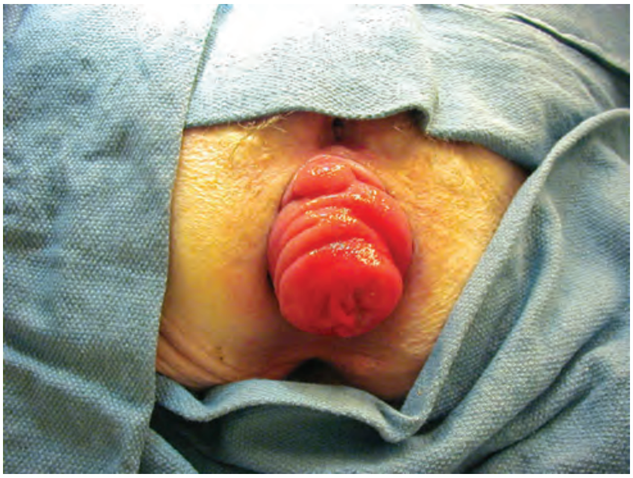


Figure 30.26 Full-thickness rectal prolapse. Note the concentric folds and palpable double wall of rectum.

The true incidence of rectal prolapse is unclear, although it is more common towards the seventh decade of life. The majority of these patients are elderly females. Patients with rectal prolapse often complain of a mass protruding from the anus, initially with straining then progressing to exteriorization with any increase in abdominal pressure and finally even at rest or simple movements. Chronic prolapse can result in an inflamed and ulcerated mucosa and bleeding. Faecal incontinence is a significant feature of full - thickness prolapse, which may be a consequence of sphincteric disruption or pudendal stretching.

The clinical assessment of rectal prolapse often requires reproduction of the prolapse. The patient may need to sit on a toilet and strain or use enemas to reproduce the event in order to fully evaluate the extent of prolapse. Full-thickness prolapse presents with concentric folds and double rectal walls on palpation, while mucosal prolapse is visualized as radial folds that do not protrude more than 3–5 cm.

Surgical management of rectal prolapse

Rectal prolapse surgery can be performed on patients in whom conservative measures have failed, and prolapse is causing significant reduction in quality of life. Surgery can be performed via the perineal route, which is less metabolically challenging, and suitable for patients who are more frail, or the abdominal route, which may be laparoscopic or open, and suitable for patients who are more physically fit.

Transperineal operations include perineal rectosigmoidectomy (Altemeier's procedure), which is a full-thickness excision and reanastomosis of the rectosigmoid, and Delorme's procedure, which is a mucosal excision and subsequent muscular plication. Neither procedure involves fixation of the rectum to a bony prominence. Both procedures are not technically demanding, and can be repeated as necessary. However, perineal procedures generally carry a recurrence rate in the region of up to 16% for perineal rectosigmoidectomy and 40% for Delorme's procedure at 5 years. Perineal procedures carry the advantage of having lower morbidity than abdominal procedures, with the absence of an abdominal incision, less adhesion formation, no intra-abdominal anastomosis and a relatively short hospitalization.

Abdominal operations include anterior resection of rectum and dorsal and ventral mesh rectopexy, plus or minus anterior resection. The fixation of rectum to sacrum may provide better results in treating rectal prolapse, and presents the lowest recurrence rate when combined with sigmoid resection (2–10%). Dorsal rectopexy frequently results in chronic constipation because of the necessary mesorectal dissection and damage to the hypogastric plexi of nerves. Ventral rectopexy has recently become popular, as it is associated with a significantly lower incidence of constipation. There have been reports of patients with ventral rectopexies requiring a posterior STARR procedure to address the angular change in the resultant rectum. The longevity of the rectopexy procedure is significant, with fewer than 5% recurrences over 5 years. The management of recurrent rectal prolapse depends on the previous procedure. Abdominal approaches should be avoided if the patient has undergone previous perineal rectosigmoidectomy, especially if sigmoid resection is contemplated, because of the potential ischaemia and necrosis of the intervening segment between anastomoses. A repeat rectosigmoidectomy can be safely performed in this patient group.

Other pathologies associated with constipation

Constipation may be caused by any stricturing lesion and carcinoma of the colon is a particular concern. Other possible causes are strictures due to diverticular disease, ischaemia and ulcerative colitis or Crohn's disease. Painful anal lesions such as fissure, abscess or prolapsed haemorrhoids may also give rise to constipation by inhibiting the wish to defecate.

Aganglionosis and/or myenteric plexus lesions

Other than drugs, constipation may be caused by malfunction of the intrinsic nervous system of the large intestine and this in turn can be subdivided into Hirschsprung disease and idiopathic megacolon and megarectum. These are dealt with as separate issues in the following sections of this chapter.

Constipation due to extracolonic factors

Extracolonic factors that may give rise to constipation include:

- illness causing immobility (e.g. myocardial infarction)
- neurological disorders
  - multiple sclerosis
  - Parkinson disease
  - autonomic neuropathy (e.g. diabetes)
  - spinal cord injury
- endocrine disorders
  - hypothyroidism
- connective tissue disorders
  - systemic sclerosis.

The most common endocrine abnormality leading to constipation is hypothyroidism, and individuals with myxoedema may develop megacolon. Hypercalcaemia may also present with constipation. There are various neurological causes of constipation. Patients with diabetes mellitus may have severe constipation owing to an associated autonomic neuropathy. These patients demonstrate loss of colonic activity after a meal and this may be reversed by neostigmine indicating intact postganglionic neurones. It is therefore worth trying prokinetic drugs in these patients. Spinal cord injury can give rise to colonic problems which depend on the site and severity of the lesion. Spinal cord transection above L2 gives rise to decreased colonic compliance and the loss of conscious control of the external sphincter function. However, defecatory reflexes are intact. This leads to constipation, but reflex emptying of the rectum can be brought about using enemas or suppositories. Injuries of sacral nerves and the cauda equina give rise to lack of control and loss of tone of the external sphincter muscles and this may lead to both constipation and incontinence. Multiple sclerosis may lead to severe constipation via mechanisms similar to spinal cord transection. Finally, Parkinson disease can be associated with constipation because of dystonia of the pelvic floor.

Psychological factors are clearly important in a proportion of patients with constipation. It may be a presenting feature of clinical depression and in patients with anorexia nervosa constipation may be a significant feature. However, there are a proportion of patients who complain bitterly of constipation but in fact when objective testing is carried out, normal colonic transit is demonstrated with normal passage of stool. If an individual perceives themselves to be constipated and yet objective evidence indicates that their defecatory habits are within the normal range, the best approach is a careful explanation of what can be regarded as normal.

Solitary rectal ulcer

The term solitary rectal ulcer is strictly a misnomer, as the lesions may be multiple and ulceration may not necessarily be present. Although any age can be affected it is most common in young adults. The macroscopic appearance is of a red thickened area of rectal mucosa usually with a shallow ulcer in the centre. Classically it is positioned on the anterior wall of the rectum opposite the puborectalis sling 5–8 cm above the anal verge. Histologically the lamina propria is replaced by collagen and there is fibromuscular replacement of the mucosa. The aetiology is unclear but it is now believed that it is due to internal rectal prolapse or intussusception which causes trauma to the rectal wall.

Clinically, all patients have difficulty in defecation which involves going to the toilet several times a day but only actually defecating once or twice. Usually there is deep-seated perineal pain and blood and mucus are often passed. On digital examination there is usually an indurated area on the anterior aspect of the rectum. The diagnosis is made by endoscopy and biopsy. If a rigid sigmoidoscopy is being used, asking the patient to strain down during the procedure may reveal the intussusception. However, the only accurate way of demonstrating an intussusception is by using defecating proctography.

Treatment can be difficult and a conservative approach should be taken in the first instance. A high-fibre diet and the use of suppositories may prevent excessive straining and alleviate the symptoms. Biofeedback may also be useful if there is obstructed defecation. Topical steroids have been recommended but there is little evidence that these have any effect. In extreme cases surgery may be indicated. If there is complete rectal prolapse then abdominal rectopexy is indicated. However, for internal intussusception the results of this procedure are poor. Thus for patients with extreme symptoms who do not have a complete rectal prolapse, rectal excision with coloanal anastomosis or even abdominoperineal resection of rectum may be indicated.

Colonic pseudo-obstruction

Acute colonic pseudo-obstruction or Ogilvie syndrome is characterized by marked dilation of the colon in the absence of mechanical obstruction. It nearly always occurs in hospitalized patients and the vast majority have an associated condition such as infection, widespread malignancy, recent surgery or trauma particularly to the spine. The underlying aetiology is not clear but it is a form of colonic dysmotility, which is the final common pathway of a variety of physiological and biochemical disturbances. As it is often associated with pelvic pathology, damage to the autonomic innervation of the distal colon has been implicated.

Clinical features

The clinical features of acute colonic pseudo-obstruction closely mimic acute large bowel obstruction. The patient often has colicky abdominal pain and progressive distension of the abdomen is the rule. Constipation is common, although not absolute, as some patients will continue to pass a small amount of flatus or liquid stool. Nausea and vomiting are also common. On examination there is massive abdominal distension and if the caecum is distended to such an extent that its viability is compromised there will be right iliac fossa tenderness. Bowel sounds are variable and can be normal or obstructive in nature. Digital examination of the rectum usually reveals an empty rectum.

Diagnosis

The diagnosis is generally made on plain abdominal radiograph, which shows the typical appearance of a distal colonic obstruction often with the cut-off in the descending colon. Measurement of the diameter of the caecum on the radiograph is important as perforation is common once this exceeds 12 cm. Unlike mechanical large bowel obstruction, the colonic haustral and mucosal pattern may be maintained and this also distinguishes it from the toxic megacolon of inflammatory bowel disease (IBD). However, the distinction between pseudo-obstruction and mechanical large bowel obstruction can be extremely difficult to make on plain abdominal radiographs and thus any patient with suspected mechanical large bowel obstruction should have this confirmed either by sigmoidoscopy or by water-soluble contrast enema before proceeding to surgery.

Complications

The main complication of pseudo-obstruction is faecal peritonitis secondary to perforation of the caecum.

Treatment

In the first instance, management consists of intravenous fluids, nasogastric aspiration and decompression with a flatus tube inserted at rigid sigmoidoscopy. Concomitantly, it is important to correct any metabolic disturbances, treat infections and stop any medications that may have an effect on colonic motility, e.g. narcotic analgesics, anticholinergic agents and calcium channel antagonists. If this does not bring about a rapid improvement, colonoscopy with decompression of the colon should be carried out. This reduces the diameter of the caecum in about 70% of cases, but in about 40% repeated colonoscopy will be required. Recurrence of the pseudo-obstruction can be decreased by placing a drainage tube into the right side of the colon at the time of the first colonoscopy.

A randomized study has shown that 2.0 mg of neostigmine given intravenously over 3–5 minutes can resolve pseudo-obstruction in a substantial proportion of patients. When the neostigmine has been given the patient should lie supine for 60 minutes and continuous ECG monitoring should be used to detect a bradycardia. If this occurs 1.0 mg of intravenous atropine should be given.

Surgery for acute colonic pseudo-obstruction is indicated if all these conservative measures fail to bring about a reduction in the size of the caecum. In the absence of signs suggesting ischaemia or perforation of the bowel, the operation of choice is a tube caecostomy through a limited right iliac fossa incision to expose the caecum. A large Foley catheter can then be used to intubate the caecum and this should be retained for approximately 3 weeks. However, if there are signs of ischaemia or perforation, a midline laparotomy should be used. If the bowel appears to be viable then a tube caecostomy will be satisfactory. However, if there is extensive necrosis a right hemicolectomy should be performed. Under most circumstances, an immediate anastomosis should be deferred and an ileostomy and mucus fistula fashioned. It should be remembered that colonic pseudo-obstruction is associated with a significant mortality, largely owing to the underlying illnesses which are associated with this condition. In patients in whom conservative treatment is successful the mortality is about 15%, whereas in those undergoing surgical intervention it is in the region of 30%.

Irritable bowel syndrome

IBS is not a precise diagnosis but rather a group of functional bowel disorders. It can be defined as abdominal discomfort or pain associated with defecation or a change in bowel habit and with an element of disordered defecation. In the West, IBS probably occurs in 15–20% of individuals, with a higher prevalence in women. The course of IBS is highly variable, but essentially it is a chronic relapsing condition.

Clinical features

The clinical features of IBS vary in number and severity. They consist of an abnormal stool frequency (more than three times a day or less than three times a week), abnormal stool consistency (hard, loose or watery), abnormal passage of stool (urgency, straining or tenesmus), the passage of mucus and abdominal bloating or distension.

Diagnosis

Because of the imprecise nature of the symptom complex in IBS, many patients with this condition will require diagnostic tests to exclude organic disease. Normally this will take the form of barium enema and sigmoidoscopy or colonoscopy. Certainly, if a patient has irritable bowel-type symptoms associated with rectal bleeding, weight loss or other suspicious symptoms, they should be thoroughly investigated. The diagnosis, however, rests on clinical features and the diagnostic criteria agreed at the multinational consensus meeting on functional gastrointestinal disorders held in Rome in 1999 are as follows.

At least 12 weeks (not necessarily consecutive) in the preceding 12 months of abdominal discomfort or pain that is associated with at least two of the following other features:

- relieved by defecation
- onset associated with change in stool frequency
- onset associated with change in stool consistency.

Treatment

The mainstay of treatment is a confident diagnosis coupled with reassurance. Other treatment strategies depend on symptom control. If constipation is a significant feature then dietary fibre should be increased by using wheat, bran or bulking agents. If diarrhoea is predominant then loperamide or diphenoxylate may be useful. For abdominal pain, anticholinergic or antispasmodic agents such as mebeverine may be of value. There is also some evidence that low-dose antidepressants may help.

Proctalgia fugax

The term proctalgia fugax means ‘fleeting rectal pain’. It is characterized by episodes of severe anal or lower rectal pain which last for a variable length of time, ranging from a few seconds up to several minutes. Although it is a recurring condition, episodes may be separated by many days, weeks or months. The prevalence in a community is not clear but may be in the region of 20%. The pain can be extremely severe and cause considerable alarm and although there are no specific diagnostic features, sigmoidoscopy should be carried out in order to exclude organic disease and to reassure the patient. The cause of the pain is probably smooth muscle spasm but the underlying cause is obscure. Treatment is usually unnecessary as the attacks are of short duration and infrequent but in patients with frequent proctalgia fugax inhalation of salbutamol may curtail episodes of pain. Clonidine or amyl nitrate has also been recommended.

Colonic diverticular disease

The term ‘diverticular disease’ describes a broad range of clinical pathology relating to the presence of symptomatic colonic diverticula. The disease presents variably, ranging from mildly symptomatic fluctuation in bowel habit and left lower quadrant abdominal pain, through haemorrhage and localized sepsis to colonic fistulation, perforation and life-threatening peritonitis. In addition, symptoms may present as an isolated occurrence, or develop into a chronic complaint. Considered a disease of Western society and of elderly patients, its incidence is rising together with associated episodes of hospitalization. Despite the significant associated healthcare burden, the aetiology, natural history and optimal management of this disease remains controversial.

The terminology applied to describe the different clinical presentations and disease stages associated with diverticular disease can be variable. ‘Diverticulosis’ refers to the presence of asymptomatic diverticula within the colon in the absence of associated inflammation. The term ‘diverticulitis’, typically prefixed as ‘acute’ on symptomatic presentation, can represent a wide spectrum of inflammatory change from localized subclinical inflammation to generalized peritonitis. This term is used interchangeably with ‘diverticular disease’, although this more commonly describes patients running a more indolent course of disease. Both diverticulitis and diverticular disease can be further described as ‘uncomplicated’ or ‘complicated’. ‘Uncomplicated’ disease refers to presentations in keeping with localized colonic inflammation, including left-sided abdominal pain, fever and leucocytosis. ‘Complicated’ diverticular disease describes disease which has progressed beyond local inflammation to one of the known sequelae of abscess formation, colonic stricture, obstruction, haemorrhage, fistula or perforation. The term *mychosis* is used in the context of diverticular disease to describe muscular shortening and thickening of the colonic wall, commonly seen at surgery.

The first noted reference to colonic diverticula as a pathological entity was made by the French surgeon Alexis Littre in the 1700s, although this was not in relation to diverticular disease *per se*. The current terminology stems from Fleischman, who used the term ‘divertikel’ in 1815. The pathology we recognize today as diverticular disease was first described by Jean Cruvehiler in 1849. Its absence from early surgical textbooks at this time suggests the disease was not as prevalent, or not as recognized, as today. Mayo reported the first surgical resection for complicated diverticulitis in 1907, advocating primary resection, although the disease remained uncommon. Although this case series promoted primary resection, staged resection consisting of drainage and stoma, interval resection and subsequent stoma reversal remained routine until the development of alternative treatment paradigms in the modern era.

Similarly, elective surgery for recurrent disease has seen a step-change in recent years. Parks’ landmark study in 1969 suggested that recurrent disease was more virulent, and more likely to require surgery, hence elective resection was frequently recommended. The original findings of this study have been challenged in recent years, and current research suggests a more benign course of recurrent disease. This, along with the not insignificant morbidity associated with elective resection, means that several of the commonly accepted treatment algorithms for diverticular disease are currently in the process of being redefined.

Diverticular anatomy

Macroscopically, colonic diverticula are acquired saccular protrusions of the colonic wall consisting of colonic mucosa. They do not contain all layers of the bowel wall, are therefore not regarded as true diverticula, and are instead termed *false, pulsion* or *pseudo* diverticula. Typically small in size, ranging from 0.5 to 1.0 cm, they occur at points of weakness where the vasa recta blood vessels penetrate the bowel wall (Figure 30.27). They most commonly protrude in four rows situated between the antimesenteric and mesenteric taenia. Colonic diverticula most frequently occur in the sigmoid and left colon and in patients with diverticulosis 90% will have sigmoid involvement. Of note, proximal right-sided colonic diverticula most frequently seen in Asian populations are true diverticula, involving all layers of the bowel wall and therefore may be unrelated phenomena.

Epidemiology and natural history

Although rarely seen in developing regions, the disease is common in the Western world and its incidence increases with age. Establishing the prevalence of diverticulosis is difficult, given that it is by definition asymptomatic. Previous estimates suggest that 40% of 60-year-olds will have developed diverticula, progressing to more than 80% of 85-year-olds.

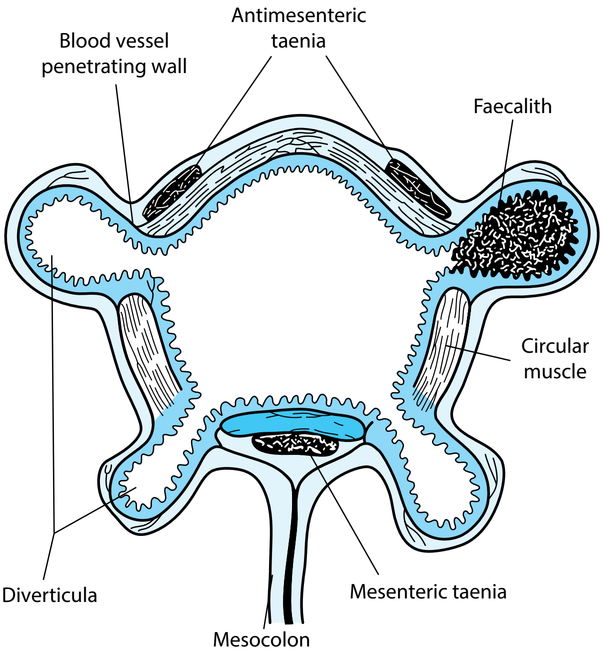


Figure 30.27 Structural abnormalities in diverticular disease.

Most patients with diverticulosis will not experience any complications related to this; however, a proportion with colonic diverticula will go on to develop symptomatic diverticular disease. The proportion affected is difficult to establish given the uncertainties surrounding the true prevalence of diverticulosis; however, it is estimated to be in the range of 10–25% with an average age of 62 years at presentation. Diverticulitis is uncommon in the young, with only 2–6% of patients affected under the age of 40 in one study. Approximately 5–15% of patients will develop bleeding complications from diverticulosis in the absence of diverticulitis (Figure 30.28). Early epidemiological studies suggested an increased male incidence; however, more recent studies now indicate a female predominance. The increased range of potential diagnoses for left - sided lower abdominal pain in women and the relatively greater proportions of elderly women than men may confound this.

Data relating to hospital admissions suggest an absolute prevalence of diverticular disease of 60/100 000 admissions in the USA. Although the disease was infrequently encountered before the twentieth century, numerous studies suggest that the incidence of symptomatic disease continues to increase in industrialized populations. In England, the NHS national admission rate increased from 0.56 in 1996 to 1.20 per 1000 population/year in 2006, with 54% aged over 70 years. In this study surgery was undertaken in 16.3% of admissions from 1996 to 2006, with 58% of cases performed on an emergency basis. The age-adjusted adult incidence of perforation has been estimated as 35 per 100 000 per annum in the UK.

Risk of recurrence has been the subject of much recent controversy, given that previously reported rates determined the policy for offering elective colonic resection early as a preventative measure. Parks' original paper in 1969 reported a readmission rate of 25% with diverticulitis. In addition, with each admission the patient was less likely to respond to medical management with increasing associated mortality. The findings of this study are now in doubt. A more recent large-scale US study

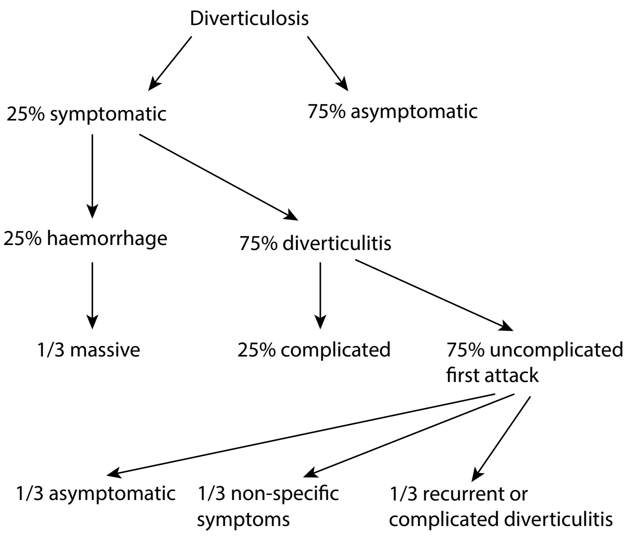


Figure 30.28 Natural history of diverticular disease. (Adapted from Fearnhead NS, Mortensen NJ, *Best Pract Res Clin Gastroenterol* 2002;**16**:577–593.)

by Broderick-Villa and colleagues suggests that the recurrence rate after hospitalization with an initial episode of diverticulitis treated medically is about 1.5% per year. Another modern large series indicated that only 3% of patients required emergency operations during 10 years of follow-up. Subsequent attacks of diverticulitis therefore appear to be similar to previous ones. Such data have led to the adoption of a more conservative treatment strategy in many centres for those hospitalized with diverticulitis.

Recent research into the role of acute diverticulitis in the development of complicated colonic diverticular disease by Humes and colleagues has suggested that most patients with complicated disease do not experience episodes of acute diverticulitis beforehand. This analysis of the UK General Practice Research Database linked to national Hospital Episode Statistics also revealed a significant excess mortality associated with complicated disease ranging from a 4.5-fold increase in those suffering from perforation or abscess to a 2.5-fold increase in those with a fistula or stricture. In this study, increasing episodes of diverticulitis were associated with an increased risk of developing a fistula (odds ratio 1.54), but there was no clear association with perforation or abscess.

Free perforation, while previously thought to be a potential complication of recurrent complicated disease, is now known to occur most frequently at a first attack of diverticulitis. In a UK study, Morris and colleagues used regional hospital coding databases and local population data to estimate the age-adjusted incidence of perforation as approximately 3.5 per 100 000 per annum. The mortality in those undergoing surgery was 23%, which compares to other studies reporting mortality ranging from 4% to 36%.

Aetiology and pathogenesis

The aetiology and pathogenesis of diverticulosis have not been conclusively proven and are potentially multifactorial. Theories relate to structural abnormalities of the colonic wall, motility or neuromuscular abnormality, and the role of dietary fibre. The most commonly cited hypothesis relates to the wide discrepancies in geographical incidence of the disease. Painter and Burkitt suggested in 1971 that the frequency with which the disease was seen in industrialized countries compared with developing nations related to the diet changes that took place in industrialized countries during the 1800s. Decreased dietary fibre, particularly through the consumption of processed food and roller milling of floor, was shown to reduce stool transit time and reduce stool weight. From this, it was proposed that resulting exaggerated colonic circular muscle contractions caused colonic wall hypertrophy as seen in sigmoid diverticular specimens. These contractions segmented the colon, increasing intraluminal pressure as confirmed on manometry studies (and as predicted by Laplace's law). This in turn could lead to the mucosal herniation through weak points in the bowel wall, as seen in diverticulosis.

Subsequent research has not demonstrated hypertrophy or hyperplasia of the taenia coli smooth muscle cells themselves, and it has been proposed that the thickening seen is caused by elastin deposition in a contracted form, leading to bunching of the taenia. Collagen abnormalities have also been implicated, with increased levels of cross-linkage in colons affected by diverticular disease. This would reduce colonic wall compliance, with resulting injury to the submucosa potentially allowing mucosal herniation to occur. Reports of diverticular disease in young patients with Ehlers–Danlos and Marfan syndrome have been cited in support of this. These structural abnormality theories remain an area of research.

Other theories relate to abnormalities of the neuromuscular apparatus in the bowel wall. Bassotti and colleagues have demonstrated that interstitial cells of Cajal and glial cells are decreased in colonic diverticular disease, potentially contributing to the disturbances in motility. Studies have investigated changes in serotonin signalling, given its importance in gastrointestinal motility. One indicated an increased presence of serotonin cells in the colonic mucosa of patients with diverticular disease. Others found attenuation of 5-HT transporter expression and function following recent acute diverticulitis.

Several studies lend support to the traditional dietary fibre hypothesis, showing a correlation between fibre intake and diverticular disease. The American Health Professionals Follow-Up Study showed an inverse correlation between dietary fibre and symptomatic diverticular disease with a relative risk of 0.58. The recent European Prospective Investigation into Cancer and Nutrition (EPIC) concluded that eating a vegetarian diet and a high intake of dietary fibre was associated with a lower risk of admission to hospital or death from diverticular disease. Vegetarians had a 31% lower risk of diverticular disease than meat eaters, confirming a risk seen with meat eaters in other studies. Despite this, there is little evidence behind the common recommendation of increasing dietary fibre intake to prevent diverticulitis; this has been extrapolated from the established correlations.

Most importantly, these hypotheses do not explain the pathogenic mechanism between the development of diverticulosis and subsequent diverticulitis. Similarities with both appendicitis and IBD have been drawn. Stasis or obstruction may lead to bacterial overgrowth and the resulting diverticula and colonic inflammation. Initial theories for this related to diverticular obstruction by a faecalith or food particle, leading to local inflammation and potential perforation. At a histological level, similarities are seen with the appearances of IBD, and several studies are currently exploring the role of aminosalicylate compounds in the treatment of diverticular disease. It has been suggested that some overlap between the diseases may exist. Additional theories include changes in colonic microbiota, ischaemia and visceral hypersensitivity which may all contribute to symptomatology.

Risk factors for diverticular disease

In addition to the risks conveyed by age, dietary fibre intake and geographical distribution of diverticular disease already highlighted, there are a number of other associated risk factors highlighted in epidemiological studies. Evidence is currently lacking as to whether these proven associations are causative and whether changing these risks will subsequently change the risk of developing diverticulitis or further attacks.

Body mass index

Body mass index (BMI) is highlighted as a risk factor in a number of studies, with an association between higher BMI and diverticulitis. Strate and colleagues showed most recently in a prospective cohort study of 47 228 male health professionals that men with a BMI ≥30 kg/m^2 had a relative risk of 1.78 for diverticulitis. A higher risk of diverticular bleeding was also noted. Rosemar and colleagues showed a linear increase in risk of diverticular disease with increasing BMI in men in a large Swedish cohort study with 28 years of follow-up, with a maximum hazard ratio of 4.4 for a BMI of ≥30.

Physical activity

Physical activity is associated with decreased rates of diverticular disease. Aldoori and colleagues showed that after adjustment for age, energy-adjusted dietary fibre and energy-adjusted total fat, overall physical activity was inversely associated with the risk of symptomatic diverticular disease with a relative risk of 0.63. The National Runners Health study revealed decreased incidence of diverticular disease of 6.2% per km/day after adjusting for age, sex and reported dietary intake.

Medication

Large cohort studies have linked a number of prescription medications with diverticular disease and its complications with non-steroidal anti-inflammatory drugs (NSAIDs). An analysis of the Health Professional Follow-up Study showed that after adjusting for risk factors aspirin use was associated with an increased risk of diverticulitis and diverticular bleeding. Users of non-aspirin NSAIDs were also at increased risk. In a case–control analysis of the UK General Practice Research Database, Humes and colleagues demonstrated a two- and threefold increase in the risk of diverticular perforation with opiate analgesics and oral corticosteroids respectively.

Smoking

Smoking has a potential association, although studies to date have provided contradictory results. Rosemar identified it as an independent risk factor (hazard ratio 1.6) in the male Swedish cohort study. In another Swedish study, the Swedish Mammography cohort, female smokers or past smokers had a higher risk of symptomatic diverticular disease (relative risk 1.89 and 1.26 respectively). However, Aldoori and colleagues did not see this association in the US Health Professionals Follow-up Study.

Classification of diverticular disease

Comparison between patients and clinical outcomes in diverticular disease is hindered by the differing presentations of uncomplicated vs complicated disease, and the varied way in which complicated disease manifests. Several classifications or scoring systems for the severity of disease have been published. The adoption of different classification systems in different publications has not always helped clarify this area.

Table 30.2 Hinchey classification and modified Hinchey classification by Sher *et al.*

Hinchey classificatio	Modified Hinchy classification (Sher *et al.*)
I Pericolic abscess of phlegmon	I Pericolic abscess
II Pelvic, intra-abdominal or retroperitoneal abscess	IIa Distant abscess amenable to percutaneous drainage IIb Complex abscess associated with fistula
III Generalized purulent peritonitis	III Generalized purulent peritonitis
IV Generalized faecal peritonitis	IV Faecal peritonitis

Table 30.3 Clinical classification of diverticulitis (adapted from Kohler *et al.*)

Grade	Clinical description	Symptoms
I	Symptomatic uncomplicated disease	Fever, crampy abdominal pain, CT evidence of diverticulitis
II	Recurrent symptomatic disease	Recurrence of above
III	Complicated disease	Haemorrhage Abscess Phlegmon Perforation Purulent and faecal peritonitis Stricture Fistula Obstruction

Most widely used internationally is the Hinchey classification, proposed in 1978 and later modified by Sher and colleagues in 1997 (Table 30.2). This is used to classify the severity of clinical insult resulting from perforated diverticular disease. Originally based on operative findings, the availability of CT imaging has extended its utility and allowed appropriate categorization of related abscess as recognized in Sher's modification. Further modification has subsequently been proposed by Wasvery and colleagues, adding a stage 0 for uncomplicated disease and introducing Ia and Ib to represent pericolic inflammation or phlegmon and pericolic abscess, respectively.

The Hinchey classification does not take into account existing conditions that may have an impact on clinical course or outcome; however, it allows an approximation at outcome based on the findings of previous studies, with risk of death <5% for stage 1 or 2 disease, ~13% for stage 3 and ~43% for stage 4 disease.

Köhler and colleagues through the European Association for Endoscopic Surgeons proposed a classification based on clinical severity and presentation, dividing this into symptomatic uncomplicated, recurrent symptomatic and complicated disease by complication. Given the subjective nature of clinical presentation for symptomatic uncomplicated and recurrent symptomatic disease, this is limited in its applicability due to the risk of incorporating an incorrect clinical diagnosis (Table 30.3).

In the European literature the Hansen–Stock classification has been widely used. This classifies the disease clinically by stage depending on its severity.

Siewert and colleagues suggested a similar classification solely for acute complicated diverticulitis based on anatomical location of an abscess (Table 30.4).

Table 30.4 Hansen–Stock and Siewert classifications of diverticular disease and diverticulitis

Hansen–Stock classification	Siewert classification
0 Diverticulosis I Acute uncomplicated diverticulitis II Acute complicated diverticulitis	
IIa Phlegmon, peridiverticulitis	I Pericolic abscess or phlegmon
IIb Abscess, sealed perforation	II Pelvic, intra-abdominal or retroperitoneal abscess
IIc Free perforation	III Free perforation
III Chronic recurrent diverticulitis	

Table 30.5 Ambrosetti's CT staging of diverticulitis (Adapted from Ambrosetti *et al.*)

Moderate diverticulitis	Severe diverticulitis
Localized sigmoid wall thickening (<5 mm)	Abscess
Inflammation of pericolic fat	Extraluminal air
	Extraluminal contrast

Ambrosetti and colleagues have described a classification system describing diverticulitis as moderate or severe according to CT criteria. Although limiting its application to patients who have therefore undergone CT imaging, it is useful in assisting decision-making by allowing patients to be categorized into optimal pathways for management (Table 30.5).

The most recent addition to this classification structure has been suggested by Klarenbeek and colleagues in 2011. This represents an attempt to combine the existing classifications into one clinically applicable system, including the most recent developments in imaging modalities and treatments (Table 30.6).

Other scoring systems may also be utilized in the setting of classifying diverticular disease such as the Peritonitis Severity Score, the Mannheim Peritonitis Score or Colorectal-POSSUM (Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity). These are not specific to diverticular disease, but are useful in the estimation of patient risk and unit outcome audits.

Clinical presentation

The majority of patients with diverticulosis remain asymptomatic. Symptomatic diverticular disease presents with a broad range of clinical manifestations, resulting from the variable nature of the disease. Symptoms may mimic those of a number of other conditions. This can take the form of acute uncomplicated,

Table 30.6 Klarenbeek classification of diverticular disease

Classification	Clinical presentation	Imaging	Treatment
A	Uncomplicated disease • Pain in left lower quadrant • Fever • Changes in relief pattern	CT scan or ultrasound • Phlegmon • Small abscess in bowel wall Colonoscopy • Diverticulosis • Inflammation	Conservative treatment • Antibiotics[a] • Low - residue diet[a] Prevention • Fibres • Prevention of obesity • Treatment of comorbidity • Mesalazine[a]
B	Chronic complicated disease • Impaired passage of stool • Presence of fistula • Recurrent rectal blood loss • Incapacitating complaints • High-risk patients	CT scan • Stenosis • Fistula Colonoscopy • Stenosis • Fistula • Blood in diverticula	Elective intervention Sigmoid resection with primary anastomosis • Open • Laparoscopically
C 1	Acute complicated disease • Fever • Painful mass	CT scan • Large abscesses (>5 cm)	Acute intervention Percutaneous drainage
2	• Ileus	CT scan • Intenstinal obstruction	Sigmoid resection with primary anastomosis Hartmann's procedure
3	• Massive rectal blood loss	CT angiography • Contrast blush colonoscopy • Active diverticular bleeding	Sigmoid resection with primary anastomosis • Open • Laparoscopically Endoscopic intervention[a] Endovascular coiling[a]
4	• Generalized peritonitis	CT scan • Pneumoperitoneum • Extraluminal contrast • Free fluid	Diagnostic laparotomy/laparoscopy • Resection with primary anastomosis • Hartmann's procedure • Lavage and drainage[a]

[a] Experimental or non - evidence - based treatement.

chronic uncomplicated or complicated disease. Thus the spectrum of differential diagnosis in diverticular disease is wide:

- urological
  - cystitis
  - pyelonephritis
  - neoplasia
- gynaecological
  - pelvic inflammatory disease
  - ectopic pregnancy
  - ruptured ovarian cyst
  - ovarian torsion
- gastrointestinal
  - colorectal neoplasm
  - ischaemic/infectious colitis
  - IBD
  - irritable bowel disease
  - appendicitis (with right-sided sigmoid colon)
- other
  - ruptured iliac artery aneurysm
  - rectus sheath haematoma.

Chronic uncomplicated diverticular disease

Chronic uncomplicated diverticular disease may present as episodic abdominal pain localized to the left lower quadrant, altered bowel habit and abdominal distension. The potential differential diagnosis is large, and the overlap with symptoms of irritable bowel disease can be significant making outpatient diagnosis difficult without further investigation. It is often diagnose when other more 'serious' pathology has been excluded. In the elderly patient group a new diagnosis of diverticular disease is frequently reached when the symptoms trigger investigation to exclude an underlying carcinoma.

Acute uncomplicated diverticular disease

Acute uncomplicated diverticular disease classically presents as abdominal pain localized to the left lower quadrant, often associated with a mild fever and leucocytosis. The picture may be variable, and right - sided abdominal pain is not infrequently seen particularly in patients with a long loop of redundant sigmoid colon lying to the right of the midline. Gastrointestinal disturbance is common, with anorexia, nausea and vomiting, constipation and/or diarrhoea frequently reported. Urinary symptoms may also occur secondary to proximity of the bladder to an inflamed sigmoid colon.

Complicated diverticular disease

Complicated disease presents according to the nature of the complication, which is predominantly abscess formation, perforation, fistulation, haemorrhage, stricture or obstruction.

Abscess formation is the most common sequela of acute complicated diverticulitis, occurring in approximately 15% of patients. Classic symptoms are in keeping with abscesses elsewhere in the body, namely spiking fever and lassitude potentially accompanied by rigors. Typical locations are pelvic and pericolic, although retroperitoneal sepsis may also develop. The severity of the infection is most commonly classified using the Hinchey classification following CT. A significant localized abscess may progress into a free perforation.

Perforation secondary to diverticular disease occurs in 1–2% of patients and is a potentially life-threatening complication associated with high morbidity and mortality. The term can describe either perforation of a diverticular abscess leading to purulent peritonitis or faeculent peritonitis from contamination of the peritoneal cavity with stool. Most commonly occurring with a first attack of diverticular disease, patients may not present with a history of diverticulitis. Presentation is similar to a generalized acute abdomen from other causes. The severity of the accompanying septic shock arising from purulent or faeculent peritonitis may help distinguish it from an upper gastrointestinal perforation and chemical peritonitis seen with, for example, a duodenal perforation.

Fistulation occurs in approximately 2% of patients with diverticular disease, and are included in the indications for surgery in 17–27% of patients. They arise from an inflamed colon and associated abscess decompressing into adjacent organs. Most often, this leads to a colovesical, colovaginal fistula, although colouterine, coloenteric and colocutaneous fistulas can occur. Often these patients do not report other associated symptoms and will initially be referred to a urologist or gynaecologist for further investigation. Symptoms are variable depending on the structures involved, although the patient is not usually acutely unwell, as any contained abscess has discharged in order to create the fistula. Colovesical fistulas account for approximately 65% of diverticular disease-related fistulas, and are suggested by recurrent urinary tract infections associated with enteric organisms, pneumaturia or faecaluria. Colovaginal fistulas account for 25% of cases and present with the passage or flatus of faecal material per vagina. Both colovesical and colovaginal fistulas are more common in women who have previously undergone hysterectomy, and it has been hypothesized that the uterus may act as a buffer, preventing the inflamed sigmoid colon reaching the bladder or vagina. Altman and colleagues reported that, compared with women who had neither hysterectomy nor diverticulitis, the risk of fistula surgery increased fourfold in hysterectomized women without diverticulitis, sevenfold in non-hysterectomized women with diverticulitis and 25-fold in hysterectomized women with diverticulitis. In all cases of fistulation, the possibility of an underlying malignancy or IBD giving rise to this should also be excluded.

Haemorrhage is not usually associated with acute diverticulitis, but usually in isolation and/or as a result of underlying diverticulosis. It most likely arises from rupture of a blood vessel involved in the herniated mucosa of a diverticulum. As the most common cause of acute lower gastrointestinal bleeding, accounting for up to 40% of cases, it is typically painless, with the only heralding symptom being the urge to defecate. Clots may be passed and the colour of the blood will be variable depending on how proximal the source is located within the colon. Unlike upper gastrointestinal bleeding, although it may initially be profuse, diverticular bleeding is typically self-limiting and resolves without need for significant intervention.

Approximately 30% of cases are associated with severe blood loss and cardiovascular compromise, although surgical intervention will be required in only 10% of these. In elderly patients, the increasing prevalence of NSAIDs and anticoagulation therapy can complicate management.

Stricture or obstruction at first presentation can be difficult to distinguish from colonic adenocarcinoma, even following radiological imaging, and the diagnosis may not be made until the lesion is resected and histologically examined (Figure 30.29). Stricturing is more common than obstruction, and is thought to develop through repeated episodes of inflammation leading to fibrosis. Patients complain of narrowed stools and constipation, varying according to the degree of structure. Approximately 10% of large bowel obstructions arise secondary to diverticular disease, and small bowel obstruction may also develop through adherence of a loop to an inflamed segment of colon or inflammatory mass. When associated with an episode of acute diverticulitis the obstruction may be functional secondary to colonic oedema and localized sepsis, and will resolve with medical treatment.

Atypical presentations may be seen in the most elderly patients or immunocompromised patients unable to mount a typical systemic inflammatory response. Abdominal pain may not reflect the severity of the clinical problem and overt signs of sepsis may be absent. Evidence suggests higher rates of abscess and free perforation in immunocompromised patients, and they are additionally more likely to fail medical management. These patients should be managed expectantly with close observation, early imaging, and a low threshold for intervention.

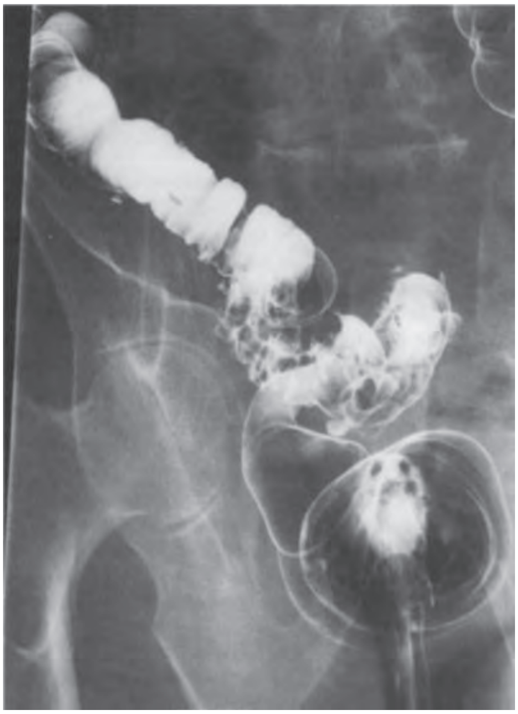


Figure 30.29 Barium enema showing a typical sigmoid diverticular stricture. Note that although the narrowing is marked, the mucosal pattern within the stricture is preserved.

Investigation

Colonoscopy

Colonoscopy is controversial in the acute setting because of the potential risk of an iatrogenic perforation of the inflamed colonic wall, or converting a sealed microperforation into a free perforation. It does however play an important role in investigating the cause of acute symptoms in the aftermath of such an attack, and is typically performed 6–8 weeks later when the inflammation has subsided. Westwood and colleagues have recently questioned this tradition of endoscopic follow-up, given the advent of routine CT on an inpatient for the majority of patients presenting as an acute hospital admission. In their study of 205 patients with acute uncomplicated diverticulitis undergoing colonoscopic investigation following diagnostic CT they showed a yield of 5.4% for advanced colonic neoplasia and 0.5% for colorectal cancer. This equated to the yield for screening asymptomatic individuals and they suggested that in this group of patients CT alone may be considered sufficiently sensitive to exclude the need for further interval endoscopic investigation in all patients. The other common indication for endoscopy is in the investigation of lower gastrointestinal bleeding. In the acute setting this may be hindered by the presence of blood within the lumen compromising the view. Endoscopy is also mandated in the investigation of stricture, when biopsies may be helpful in diagnosing the aetiology, and in the investigation of suspected diverticular fistula (Figure 30.30).

Ultrasound

Ultrasonography is frequently used in place of CT in European centres. With a reported sensitivity of 84–98% and a specificity of 80–97% in studies, it useful non-invasive, non-ionizing investigation on which acutely inflamed diverticula can be demonstrated. Abdominal ultrasound also has an important role in identifying other potential causes of acute symptoms, particularly relating to gynaecological causes in female patients. It can also be utilized to monitor a known abscess or mass. Uptake of ultrasonography for diagnosis of diverticular disease has not been universal however due to the operator-dependent variability and localized imaging window provided, which may

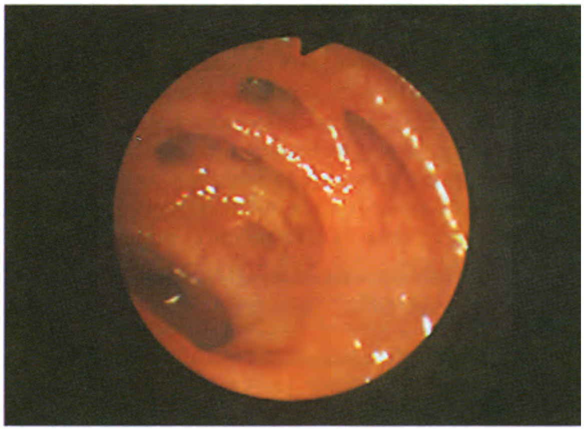


Figure 30.30 Colonoscopic appearances of diverticular disease. Note that in severe disease such as this it can be difficult to distinguish between the lumen and the mouths of the diverticula.

miss other important related findings elsewhere in the abdomen or pelvis.

Computed tomography

CT is considered the primary diagnostic investigation in the acute setting by many, and diagnoses acute diverticulitis with a sensitivity approaching 100% and a specificity of 95%. The main benefit surrounds its ability to identify alternative pathology in the acute setting. In addition, it provides the best modality for characterizing extramural disease and abscess, which have led to a number of diverticular disease classifications based on CT findings as discussed earlier in this chapter. This categorizing of patients can be useful in guiding subsequent clinical management. The most frequent findings consistent with acute diverticular disease are bowel wall thickening, fat stranding and diverticula. Rectal contrast may also be used to help delineate suspected fistula tracts, or demonstrate perforation. CT is also a useful therapeutic tool for percutaneous drainage of intra-abdominal abscesses. The main concerns relate to exposure to ionizing radiation, particularly in younger patients.

Contrast studies

Contrast studies were historically used routinely prior to the advent of widespread access to CT for initial investigation of suspected diverticular disease. Water-soluble single-contrast enema has now largely been superseded by CT, although it remains an option in resource-poor environments. It may help identify intramural changes and leakage of contrast may aid diagnosis of perforation or fistula; however, it does not provide information relating to extramural disease.

Further investigations

Other investigations may also be required in complicated disease, depending on the specific presentation and concerns. In significant lower gastrointestinal bleeding, mesenteric angiography may be used in order to localize the bleeding point within the colon. This may then be treated radiologically or guide surgical resection. In suspected diverticular fistulation, cystoscopy may be useful in identifying vesical openings and ruling out other urological disease causing symptoms.

Treatment

The optimal treatment of diverticular disease is an area in evolution and the subject of active research. Traditional management recommendations are being questioned, with an increasing move towards conservatism. New medical therapies are being investigated, and new paradigms in surgical management are being proposed.

Medical treatment

For mild, uncomplicated diverticulitis outpatient management with or without antibiotics may be appropriate, depending on the patient's individual situation. Further investigation may be undertaken if warranted after the episode has resolved. If outpatient management fails then hospitalization is required for more active treatment. For patients presenting with more acute systemic symptoms including fever, pain and inability to tolerate oral intake then hospitalization is appropriate from the outset. Supportive therapy, including antimicrobial therapy, gut rest, intravenous fluids and analgesia may be required together with radiological imaging to confirm diagnosis and stratify treatment. A proportion of patients will fail medical therapy, and approximately 15% of patients develop pericolonic or intramesenteric abscess. In these cases radiological or surgical intervention may ultimately be required.

Antibiotics remain the main therapy for symptomatic diverticular disease. The precise regimen will depend on local recommendations; however, typically these will be chosen to cover both aerobes and Gram-negative micro-organisms. In cases of severe sepsis these may be given parenterally until the sepsis is controlled, followed by oral administration. Guidance on the optimum duration of oral antibiotic therapy is lacking, but this will typically be prescribed for 7–10 days.

Currently the role of antibiotics in mild uncomplicated diverticular disease is being questioned. In Europe there is a trend towards avoiding antibiotic therapy in these cases, whereas in North America antibiotics are more frequently administered. A Scandinavian study has suggested bowel rest alone in these cases is similar in efficacy to antibiotic therapy. This area is currently being examined further in the DIABOLO trial, a large multicentre randomized control trial in Europe which is investigating a liberal policy, without antibiotics and without the strict requirement for hospital admission. The clinical and economic results are awaited.

Risk factor modification is a traditional component of medical therapy based on demonstrated associations, although the evidence that modifying these risk factors imparts a beneficial effect is largely lacking. Increasing dietary fibre, weight loss, smoking cessation and regular physical activity may play a role in reducing the frequency or severity of the disease.

Emerging medical therapies include the use of aminosalicylates and probiotics, and active studies continue to investigate their role. Low-grade proinflammatory states are associated with active diverticular disease; therefore, immunomodulatory agents such as 5-aminosalicylic acid (5-ASA) compounds used in the treatment of IBD have garnered interest as potential agents for treating diverticular disease or preventing symptomatic attacks. A meta-analysis of several small trials investigating 5-ASA products in the treatment of diverticulitis has confirmed improved outcomes; however, the results must be interpreted cautiously until larger objective trials have been conducted. Probiotics have been investigated as a similar potential treatment for altering the inflammatory milieu seen in diverticular disease. It has been theorized that the altered bowel motility and transit time may change the colonic microflora. Probiotics are live micro-organisms, the ingestion of which may restore normal colonic microflora. The beneficial effect of this treatment is yet to be clearly demonstrated.

Radiological intervention

If systemic symptoms and signs of sepsis continue despite optimum antimicrobial therapy, the presence of a diverticular abscess should be suspected. Following confirmatory CT imaging, percutaneous radiologically guided abscess drainage may bring the sepsis under sufficient control for antimicrobial therapy to succeed without further surgical intervention. Despite this, radiological drainage may fail and approximately 20–30% of cases will ultimately require surgical drainage.

The options for radiological intervention depend on the size, nature and anatomical location of the abscess(es). A single, small, well-contained abscess <5 cm in size may ultimately respond to antimicrobial therapy alone. Percutaneous drainage tends to be reserved for an abscess ≥5 cm in size. Multiple or multiloculated abscesses may still be drained; however, the failure rate is higher. Pelvic collections may be difficult to drain percutaneously, although transrectal, transgluteal or transvaginal drainage is an option. Typically a drain will be left in the collection until drainage is minimal, and samples of the fluid drained should be sent for appropriate microbiological investigations to ensure optimum antimicrobial therapy. Radiological intervention is also appropriate for patients with lower gastrointestinal bleeding secondary to diverticulosis. Most cases spontaneously settle with conservative management, although acute ongoing or recurrent massive bleeding may require further intervention.

In cases posing a continuing problem investigation can be undertaken with tagged red cell nuclear scans. More sensitive than mesenteric angiography, these can detect bleeding rates as low as 0.1–0.5 mL/min, although at a lower resolution. Nonetheless, undetectable bleeding obviates the need for invasive mesenteric angiography. Radionuclide suffers from a high rate of false localization due to tagged red cell migration in the bowel and overlying bowel loops. This ranges from 14% to 25% in reported studies, hence directed segmental resection for bleeding should not be performed on this basis alone.

In acute cases, and where the facilities and appropriately trained radiologists are available, the first - line treatment is mesenteric angiography to identify the bleeding point, with selective embolization of the feeding vessel to control this. One large Australian series reported a success rate of 76% associated with embolization, but in order to do this active bleeding must be present to allow the bleeding point to be located. Bleeding rates of 1–3 mL/min are the minimum that can be identified radiologically. In this series it was seen in only 48% of patients, and amenable to embolization in 45%, although lack of active bleeding is clinically reassuring. This procedure is not without risk, and cases of ischaemic bowel, false aneurysm and sepsis are reported, along with the risk of further major bleeding.

Surgical treatment

Surgical intervention is reserved for cases in which medical management has failed, or for the complications of diverticular disease. Although the majority of cases of diverticulitis will resolve with medical management, approximately 20% of patients with diverticulitis require surgical treatment at some stage. In recent years there has been an increasing trend towards more conservative management options, and considerable debate still surrounds the need for operative intervention, the timing of this surgery, and the appropriate procedure to perform.

Elective surgery

Recurrent diverticulitis has previously been an accepted indicator for elective sigmoid resection, for both symptom control and the prevention of recurrence and more complicated disease. Widely endorsed guidelines, until recently, recommended surgery for any patient who had experienced two attacks of acute diverticulitis requiring hospitalization. Such elective resection in diverticular disease carries not insignificant risks, with morbidity estimated at 25–50% and a mortality of 1–15%, increasing with age. This was justified by historical studies suggesting high rates of recurrence with hospitalization and complicated disease after repeated attacks of diverticulitis. Contemporary studies have challenged this justification, suggesting that the disease in fact follows a more indolent path. The current trend is therefore moving towards more conservative management, and this policy has not so far led to an increase in patients presenting with complicated disease. In addition, there has been a growing recognition that such elective surgery does not always improve the patients' symptomatology and disease recurrence may still occur in 3–13%.

Elective surgery should now only be offered following a full consideration of factors including the patient's age, comorbidities, severity of disease, complications and risk of further episodes. Special cases include patients in remote locations who may have difficulty in accessing medical care for the complications of diverticular disease when they occur, and those in whom repeated hospitalization is unacceptable for work or lifestyle reasons.

Diverticular fistulas may occasionally close spontaneously when the acute inflammatory process settles. More often those not resolving with conservative management require surgical treatment unless patient choice or operative risk prevents this. Patients and clinicians must be aware of the small risk of underlying neoplasia driving the fistulating process rather than a pure diverticular phenomenon. Often this can only be established on histological examination of the excised specimen. The operation should be timed to ensure the patient is nutritionally optimized and inflammation is controlled or quiescent. An attempt at establishing the location of the fistula should be undertaken prior to surgery through appropriate contrast imaging or endoscopic examination. Despite this, at surgery the tract may not be clearly identified and resection of the involved portion of colon may be sufficient to resolve the problem. Primary anastomosis will usually be performed, and the omentum may be mobilized and placed between the bowel and other pelvic organs. The surgery may be technically difficult due to the significant inflammatory mass often associated with such cases, and anatomy may be distorted. Placement of ureteric stents may be considered prior to surgery. Although case series report success undertaking such procedures laparoscopically, their complexity makes this a difficult undertaking.

Emergency surgery

Haemorrhage secondary to diverticulosis is an infrequent cause for emergency surgery, and rarely with symptomatic diverticulitis. As discussed earlier in this chapter, although approximately one-third of patients with bleeding experience severe blood loss this will largely spontaneously settle with only supportive therapy. Only 10–20% of patients will require further intervention. This most commonly takes the form of interventional radiological treatment with mesenteric angiography and embolization of the vessel feeding the bleeding point as discussed earlier in this chapter. Rarely, patients will nonetheless require surgery due to massive recurrent or ongoing bleeding, or failed mesenteric angiography.

The need for surgery in this patient group is related to the ongoing transfusion requirement, with one study indicating that in those receiving ≥4 units of blood 60% will ultimately require surgery. With ≤4 units only 1.5% required surgery. Patients coming to surgery are typically haemodynamically unstable, high-risk cases.

Once an upper gastrointestinal cause is clearly excluded with an oesophagogastroduodenoscopy, laparotomy and on – table colonoscopy may be attempted. The colon is lavaged through the appendix stump in order to clear retained blood and improve the endoscopic view. This may be unrewarding; however, if a bleeding point is clearly identified then a directed segmental colonic resection can be performed with anastomosis or exteriorization of the bowel. If no bleeding point is seen then blind subtotal colectomy can be performed with end ileostomy and mucus fistula. In this situation the patient is often *in extemis* and the surgery has substantial morbidity and mortality.

Stricture or bowel obstruction may result from fibrosis and scarring caused by recurrent episodes of inflammation. Stricturing is seen more frequently than obstruction, which is usually incomplete. Management of these cases depends on the degree of obstruction and urgency with which it presents. With strictures and incomplete obstruction, bowel rest and treatment of the underlying inflammation may be all that is required to get a patient over an acute episode. Full outpatient radiological and endoscopic investigation may then be performed. In cases of complete obstruction, emergency management will be required. In both scenarios, colonic stenting can be considered, either as a temporizing measure and bridge to surgery or as a non-operative option. The procedure is technically more difficult than that for colorectal cancer, because of the longer portion of bowel typically involved. The risk of stent migration is high, and series published so far indicate that complications and need for reintervention are both common. Surgical options include on – table colonic lavage, resection and primary anastomosis, with or without covering ileostomy or alternatively the traditional Hartmann's procedure. Occasionally, small bowel obstruction may develop through its adherence to or involvement in an inflammatory mass. Although CT scanning may help differentiate between primary small bowel obstruction and a secondary cause relating to diverticular disease, exploratory surgery may still be required.

Perforated diverticular disease has traditionally mandated surgery unless clinically futile. Classic dogma taught that Hinchey stage I and potentially stage II diverticular perforations are treated surgically with resection and immediate anastomosis. Resection and diversion via a Hartmann's procedure were reserved for perforated stage III and IV disease. This was based on the theory that an anastomosis in the setting of local infection and/or systemic inflammation is associated with a high risk of anastomotic leak. Evidence is now growing that this may be overtreatment and that such a radical approach may be avoided in many cases. The theory for this comes from the concept of microperforation of the bowel, which subsequently seals. Purulent contamination may therefore not be in direct continuity with the lumen, and may arise from a ruptured interloop abscess.

For Hinchey II and III patients, studies suggest that laparoscopy with copious peritoneal washout with continued antimicrobial therapy may be better treatment provided faecal contamination or free perforation is not seen. If these are seen, conversion to open surgery is still recommended. One prospective multi-institutional study has shown laparoscopic washout is associated with significantly improved outcomes compared with standard laparotomy, with morbidity of 4% and mortality of 3%. Such treatment remains controversial in some quarters, and randomized controlled trials are under way. The surgical options in perforated diverticular disease can be summarized as follows:

- conservative
  - laparoscopic lavage and drainage
  - laparotomy ± suture, drainage and stoma formation
  - exteriorization of sigmoid loop
- radical
  - resection and diversion without anastomosis (Hartmann's procedure)
  - resection and primary anastomosis
  - resection and primary anastomosis with proximal diversionary stoma.

Adapted from ZH Krukowski, in Philips RKS (ed.) *Diverticular Disease* London,UK:Saunders,2010.

When a laparotomy is required, the sigmoid colon should be resected in order to remove the inflammatory focus, rather than simple drainage and washout. A one-stage procedure is preferable in order to avoid the mortality and morbidity associated with further surgery; however, careful patient selection is required for this. A primary anastomosis should only be performed in an adequately resuscitated, cardiovascularly stable patient and where appropriate high-level perioperative care is available. Splenic mobilization should be considered in order to reduce tension on the join. When concerns arise, defaulting to a stoma may be the safest option.

The preference for one-stage surgery has developed partly through increasing experience in performing anastomoses in what would previously have been considered adverse conditions, partly through better patient selection and partly through improved anaesthesia and perioperative care. In parallel with this, studies have shown the high levels of morbidity and mortality associated with a Hartmann's procedure. Stoma complications of over 10% are reported in this setting, and given the elderly population typically affected up to 35% of patients may never have their colostomy reversed due to the risks of further surgery or patient preference. A recent systematic review reported mortality of 9.6% vs 15.1% for primary anastomosis versus Hartmann's resection for perforated diverticular disease.

In this population it is also important to ensure sigmoidoscopy is performed prior to resectional surgery in order to exclude coexistent distal adenocarcinoma, as this may influence the surgery. Sigmoidoscopy can be quickly performed on the operating table so as to avoid any delay to the emergency surgery.

In summary, our understanding of diverticular disease and the optimum treatments for this are in a state of transition. Although diverticular disease is becoming more common, its natural history is still being defined. Several theories regarding the aetiology of the condition exist, and numerous epidemiological risk factors have been identified, however the underlying pathogenesis has not yet been proved. New medical treatments are being investigated, and classically held views regarding the optimum surgical management are being challenged by new paradigms.

Ulcerative colitis

Ulcerative colitis is a distinctive form of inflammatory disease affecting the large intestine. Inflammation of the colon (colitis) has varied aetiology but the various disorders can be grouped as infective or non-infective:

- Infective
  - Virus
    - Cytomegalovirus (CMV)
  - Bacteria
    - *Campylobacter*
    - *Escherichia coli*
    - *Shigella*
    - *Clostridium difficile*
    - *Chlamydia*
    - *Gonococcus*
  - Protozoa
  - Amoebiasis
  - *Cryptosporidium**
  - *Giardia**
- Non-infective
  - Ulcerative colitis
  - Crohn's disease
  - Radiation enteritis
  - Drug-induced colitis

***Especially in immunocompromised patients.**

Ulcerative colitis is a disease confined to the large bowel mucosa. The disease chiefly affects the young and has an equal sex distribution. The annual incidence per 100 000 population is similar for males and females up to the fourth decade of life. As a rule of thumb, it is useful to remember incidences of 10 and 5 per 100 000 for ulcerative colitis and Crohn's disease respectively. The incidence of ulcerative colitis has changed little over the last 30 years, whereas Crohn's disease has increased about fivefold, although this may now be stabilizing and is possibly in decline. The prevalence of ulcerative colitis is about 160 per 100 000 population (compared with about 50 per 100 000 for Crohn's disease). This means that there are around 100 000 people affected in the UK.

Epidemiology and aetiology

Genetics

Between 10% and 20% of affected individuals have a first-degree relative with IBD. There is a concordance within Jewish families, a low incidence in spouses and absence of IBD in families of adopted probands. Crohn's disease and ulcerative colitis can occur in the same family, and the overlap of features of the two diseases (indeterminate colitis) of 10–15% and the change of diagnosis from one to the other in a further 10% may be a feature of genetic heterogeneity. While it has been suggested that both diseases share some similar gene loci with other genes defining each condition, it is noteworthy that there are no reports of mixed Crohn's disease and ulcerative colitis among monozygotic twins.

Extra-alimentary manifestations, including ankylosing spondylitis and primary sclerosing cholangitis, are more common in first-degree relatives of affected *propositi*; both have HLA associations, including HLA-B27 and HLA-B8 respectively. Ulcerative colitis is more common in white people than in black people or Arabs. Whether this is genetic or environmental is unknown. IBD has a low incidence in developing countries.

Environmental factors

The incidence of ulcerative colitis rises about 10 years ahead of Crohn's disease, suggesting environmental influences. There is evidence that NSAIDs can be associated with IBD in humans. It has long been recognized that some patients with ulcerative colitis have a history of infective proctocolitis. Smoking is protective in ulcerative colitis but not in Crohn's disease. It also appears to be true for pouchitis. Data are conflicting on the influence of oral contraceptives, and most of the information relates to Crohn's disease. No causative dietary factor has been identified in humans. However, lactose intolerance can accompany ulcerative colitis, although this is rare.

Pathogenesis

Inflammation is confined to the large intestine, which includes the colon, rectum and upper anal canal. The mucosal columnar glandular epithelium extends into the anal canal to the anal transitional zone, which varies in longitudinal length from a few millimetres to over a centimetre. The anatomical extent of ulcerative colitis varies from involvement of the upper anal canal and rectum alone (proctitis) to the colon more proximally (proctocolitis). The rectum is always involved for all practical purposes, although relative rectal sparing can occur in patients receiving local anti-inflammatory treatment. A spared rectum not associated with local treatment should raise the suspicion of Crohn's disease. Backwash ileitis occurs only in cases with colonic extension to the ileocaecal junction. Anal disease occurs in about 10% of cases coming to proctocolectomy. The lesion is usually minor, e.g. a low fistula or fissure. Rectovaginal fistula can occasionally occur in ulcerative colitis.

The inflammation in the colon and rectum is diffuse without intervening normal mucosa. Ulceration causes bleeding and in patients with severe disease the inflammatory exudate results in

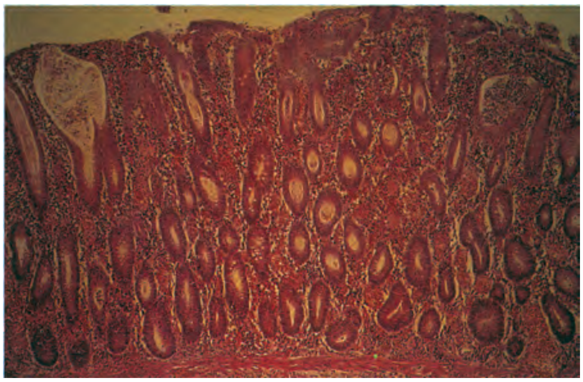


Figure 30.31 Histological appearances of ulcerative colitis. Note the crypt abscesses.

loss of water, electrolyte and protein which may be as great as 200 g per 24 hours (Figure 30.31).

Clinical presentation

At presentation, approximately 50% of patients have disease confined to the rectum (proctitis). In 30% this extends to the left colon (proctosigmoiditis) and in a further 20% disease extends beyond the splenic flexure (extensive colitis). Symptoms are local and general. The severity of the former and the presence of the latter depend largely on the anatomical extent of the disease. Ulcerative colitis is characterized by exacerbations and remissions. Bloody diarrhoea with urgency is the hallmark of colitis.

Proctitis

Symptoms include bleeding and mucus secretion. Sometimes constipation occurs but more often there is increased frequency of defecation. Rectal irritability may result in urgency of defecation. Systemic symptoms are very uncommon. Patients do not suffer from disturbances of growth, and only rarely from extra-alimentary manifestations or subsequent cancer. There is a tendency for proctitis to extend proximally with time.

Proctosigmoiditis and extensive colitis

Proximal extension to the left colon and more proximally leads to worsening local symptoms and systemic disturbances in some cases. Urgency is the most incapacitating local symptom. When severe, patients may have warning of impending defecation of a few seconds only. In such cases, urge incontinence can occur. Severe symptoms often dominate the patient's life and seriously affect work and family life. The protein-losing enteropathy may lead to malnutrition with loss of lean body mass and anaemia. Retardation of growth in children may be a feature of extensive colitis. In acutely ill patients, water and electrolyte loss may cause hypovolaemia and breakdown of the mucosal barrier may lead to toxicity.

Exacerbations may be precipitated by anxiety or stress but usually there is no recognizable causative factor. The disease may be of the acute relapsing type, with acute episodes interspersed by periods of complete resolution. Alternatively it may take the form of persisting chronic disease. Such patients may develop acute exacerbations that settle only partially on treatment. Patients with extensive disease are more likely to have associated extra-alimentary manifestations and are at greater risk of developing malignancy. These complications can occur in patients with disease confined to the left side of the colon but are much more frequent in extensive colitis.

Acute presentation

About 5% of patients present with acute severe colitis as the first manifestation of the disease. The patient will be ill with severe local symptoms, weight loss, anorexia, and water and sodium depletion. Intensive medical treatment has a high chance (70%) of inducing remission but when unsuccessful urgent or semi-urgent surgery will be necessary. Acute severe colitis may progress to toxic dilatation recognized by distension of the colon to a diameter greater than 6 cm on a plain radiograph. Perforation is a rare but serious occurrence with a mortality still approaching 40%. Rarely deep ulceration (usually in the rectum) may cause severe bleeding.

Extra-alimentary manifestations of ulcerative colitis

Up to one-third of patients with ulcerative colitis will develop at least one extra-alimentary manifestation during the course of the illness. These can be divided into those related or not to disease activity. Amyloid or hypertrophic osteoarthropathy are rare and are the result long-standing chronic illness.

Arthropathy

Arthropathy is the commonest extra-alimentary manifestation. It can be divided into three broad groups.

Activity - related polyarthropathy occurs in up to 20% of patients and is more likely in those with extensive disease. It affects predominantly the large joints of the limbs, knees being the most common. The arthropathy is fleeting and asymmetrical and is rheumatoid factor negative. It disappears when medical treatment induces a remission or after proctocolectomy. It can develop in patients with pouchitis after restorative proctocolectomy.

Ankylosisng spondylitis is an axial arthropathy involving the sacroiliac joints and one or more vertebrae and occurs in up to 5% of patients. The majority of such cases are HLA-B27 positive. The disease is unrelated to the activity of colitis and does not respond to proctocolectomy. There may be a genetic basis.

Asymptomatic sacroileitis is an arthropathy limited to the sacroiliac joint and is HLA-B27 negative. It occurs more frequently than ankylosing spondylitis and is also unaffected by treatment for the colitis.

Liver

Ulcerative colitis - associated hepatic and extrahepatic disorders occur in up to 5% of cases, predominantly in those with extensive colonic involvement. Fatty degeneration is common, but has no obvious clinical importance. Parenchymal liver disease of the chronic active hepatitis type and cirrhosis can occur. The latter may lead to portal hypertension.

Primary sclerosing cholangitis is more often seen in ulcerative colitis than in Crohn's colitis. The disease is characterized by a fibrous inflammatory reaction within the biliary tree leading to multiple intra- and extrahepatic stenoses. The diagnosis is made on endoscopic retrograde cholangiopancreatography or MRI. There is no apparent relationship between duration of disease and disease activity, although patients with primary sclerosing cholangitis undergoing restorative proctocolectomy have a higher subsequent incidence of pouchitis and dysplasia in the ileal pouch mucosa. Treatment by steroids, colectomy or antibiotics is ineffectual and ultimately the disease progresses to liver failure. Such patients may be considered for liver transplantation.

Cholangiocarcinoma is a rare association with ulcerative colitis. There may be an induction period of many years and the risk appears to continue even after proctocolectomy.

Skin

Erythema nodosum is the commonest cutaneous manifestation of IBD. It occurs more often in Crohn's disease. The condition is activity related. Pyoderma gangrenosum is more often associated with ulcerative colitis. It usually occurs in the lower limb as a circumscribed area of erythema with a punched-out ulcerated centre. Lesions may be multiple and occasionally are very extensive. Proctocolectomy is associated with healing in about 50% although this may take weeks to months.

Eyes

Uveitis is rare and is not related to disease activity. The condition can lead to scarring with visual impairment and ophthalmological management is essential. Episcleritis is activity related and occurs more often in Crohn's disease. It does not lead to chronic changes.

Cancer

The occurrence of malignant transformation has been known for years but it was not until 1967 that dysplasia was recognized as a histopathological marker for impending or actual malignancy. Ulcerative colitis should therefore always be considered when large bowel cancer presents at an early age. The incidence of cancer depends on the duration of the disease. This is estimated to be less than 1% within 10 years of onset, increasing to 10–15% in the second decade and to over 20% in the third. As a rule of thumb after 10 years of colitis the incidence of colorectal cancer increases by 1% per year (Figure 30.32).

Colonoscopic surveillance relies on the identification of flat dysplasia or a dyspasia - associated lesion or mass. There is general agreement among pathologists on the criteria for its diagnosis. The presence of low-grade dysplasia is as likely as high-grade dysplasia (54% vs 67%) to be associated with an already established cancer.

Cancers can be missed by colonoscopic surveillance programmes. The American College of Gastroenterology Practice Parameters Committee has recommended colonoscopy annually or every 2 years with multiple biopsies taken at 10 cm intervals through the colon and rectum. Review of the blopsies by an experienced histopathologist is essential.

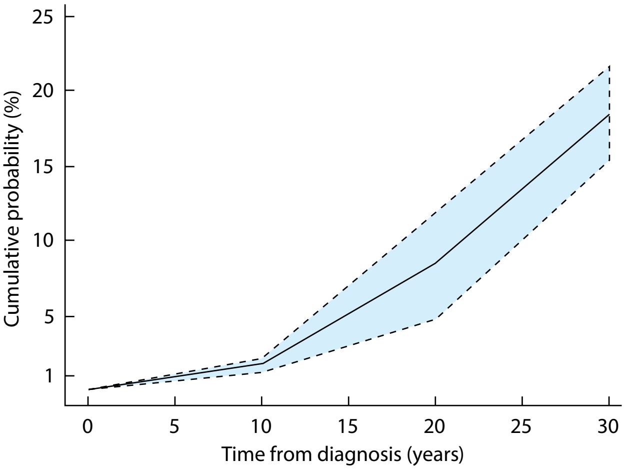


Figure 30.32 Risk of developing cancer in ulcerative colitis among 194 studies of 54, 478 patients. (Adapted from Eaden *et al.*, Gut 2001; 48:526–35.)

Investigation

In the tropics, infective causes constitute the vast majority of causes of colitis. In temperate regions infective causes may occur in hospitals and long-stay institutions. The diagnosis is made by histopathological examination of biopsy material taken during endoscopy having excluded microbiological causes. Endoscopy will determine the extent of inflammation in the assessment of severity.

Microbiology

A specimen of stool must be sent for microbiological examination. If amoebiasis is suspected, the specimen should be examined in the laboratory within a few hours. In addition, diagnosis of amoebiasis requires a biopsy to demonstrate cysts. *Shigella, C. difficile* and *Campylobacter* infection should be excluded. These may occur in epidemics in institutions with a significant mortality in frail elderly patients. The microbiologist should be warned on the request form that these could be present.

Proctitis can be caused by gonorrhoeal and chlamydial infection. The inflammation is catarrhal and consists of an erythematous flare associated with a purulent exudate. It rarely extends proximally beyond a few centimetres from the anal verge. When suspected, rectal, urethral and vaginal swabs should be taken. Again the microbiologist should be forewarned. Opportunistic infection may cause proctocolitis in immunocompromised patients (e.g. those with human immunodeficiency virus infection) or those on immunosuppressive drugs. Examples include cytomegalovirus, *Mycobacterium avium-intracellulare* and cryptosporidia.

Endoscopy

Loss of the vascular pattern (the submucosal vessels seen through the transparent mucosa) is the most sensitive sign of inflammation. This is due to oedema of the mucosa which makes it opaque. Oedema also causes fine granularity in which there is a delicate regular stippled appearance of the mucosal surface. More severe changes include erythema, contact bleeding and frank ulceration. Where previous acute attacks have been followed by repair, mucosal regeneration nodules (coarse granularity) or pseudopolyps may be seen. Pseudopolyps represent tags of mucosa that have been partially detached during the active episode and remain as projections after healing of ulceration. Rigid sigmoidoscopy will only visualize the rectum. Colonoscopy allows assessment of the proximal extent of the disease.

Histopathology

Histopathological examination of a mucosal biopsy is the basis of diagnosis.

Biopsy technique

A biopsy is obligatory and is most easily obtained during rigid sigmoidoscopy. Colonoscopy used for surveillance allows multiple biopsies to be taken. Perforation and bleeding are potential complications. The patient must be asked whether anticoagulants or immunosuppressive drugs are being taken before a biopsy is performed. The biopsy taken during rigid rectoscopy itself should be obtained with forceps with a circular cusp that minimizes the depth of penetration. The optimal site is about 7 cm from the anal verge in the posterior quadrant of the rectum. Adequate vision during rectoscopy must be assured and the jaws of the forceps are firmly closed, taking a bite of mucosa and submucosa. After the biopsy has been taken, the site must be inspected for bleeding. If this persists, a topical solution of epinephrine (adrenaline) 1 in 1000 soaked in a small swab should be applied to the biopsy site. The biopsy should be oriented onto a piece of absorbent paper and placed in formalin (10%).

Histopathological features

Active disease

In active disease, there is mucosal thickening with infiltration of the lamina propria by neutrophils, plasma cells, lymphocytes, eosinophils and mast cells. Mucin within goblet cells is discharged so that these are less evident or absent (goblet cell depletion). The degree of neutrophil infiltration is the best histopathological marker of severity. In mild disease, neutrophils are confined to the lamina propria. Extrusion of neutrophils into the crypt lumen forms a crypt abscess, the number of which correlates with the severity of disease. Mucosal ulceration is partly the result of rupture of crypt abscesses leading to mucosal destruction. Damage to the crypt basal epithelium leads to loss of crypts. Attempts at regeneration may be mistaken for dysplasia but the presence of more normal cells towards the luminal surface allows these to be distinguished. There may be branching of crypts owing to regeneration following crypt epithelial damage.

Acute severe colitis

Progression of these acute changes occurs in cases with acute severe colitis. Ulceration can be very extensive, leaving large areas of exposed muscularis propria covered with granulation tissue. This may be associated with thinning of the musculature and colonic dilatation. Inflammation may be transmural and fissure formation may be seen. Colitis in remission may leave a distorted architectural pattern with crypt depletion. Mucosal cells that remain often regain normal function and show retained mucin as identifiable goblet cells. A chronic inflammatory cell exudate in the lamina propria is likely to be present, although this may be very mild in patients in remission for long periods. Paneth cell metaplasia indicates episodes of previous colitis.

Table 30.7 Histopathological distinction between ulcerative colitis and Crohn's disease

	Ulcerative colitis	**Crohn's disease**
Macroscopic		
Distribution	Colon and rectum	Gastrointestinal tract
Rectum	Involved	Often spared
Anal disease	Rare	Common
Malignant risk	10% at 20 years	Probably similar in large bowel disease
Intestinal fistula	Never	Common
Stricture (non-neoplastic)	Rare	Common
Microscopic		
Bowel wall involvement	Mucosa and submucosa	Full thickness
Granulomas	None	60–70%
Mucus secretion	Impaired (goblet cell depletion)	Slightly impaired
Fissuring	Absent	Common
Crypt abscess	Common	Rare

Ulcerative colitis/Crohn's disease

Diagnostic difficulties in differentiating ulcerative colitis and Crohn's disease have been recognized for many years. The pathological criteria distinguishing them are shown in Table 30.7.

Indeterminate colitis

In some patients insufficient numbers of these diagnostic attributes are present or there is considerable overlap and atypical features are seen. Thus it may be impossible for the pathologist to separate the two diseases, which may be reported as unclassified colitis or as unclassified colitis with additional indication of the possible or probable presence of Crohn's disease or ulcerative colitis. In about 10% of cases, however, the pathologist will be able to state only that the colitis is indeterminate.

Indeterminate colitis is not a disease entity. It is a term which indicates that the histopathologist is unable to come to a firm diagnosis owing to the presence of features of both conditions. Usually the dilemma arises in emergency colectomy specimens where severe inflammation may be combined with features of ulcerative colitis and Crohn's disease.

In trying to resolve the diagnostic dilemma more biopsies should be taken. If the patient has had a colectomy for acute disease, then these will come from the rectal stump which may have already developed diversion proctitis making the histopathologist's task more difficult. The small bowel should be examined by endoscopy or radiology which if abnormal is suggestive of Crohn's disease as would be the presence of an anal lesion.

When histopathological, radiological and clinical features are considered together, patients with indeterminate colitis can usually be judged to incline more to Crohn's disease or ulcerative colitis. Where they cannot, the natural history tends to incline to that of ulcerative colitis.

Radiology

A plain abdominal radiograph is the most useful means of identifying colonic dilatation. There has been a movement away from tubular contrast radiology to CT. The instant barium enema is now rarely used although it gives an excellent record of the extent of the disease in most cases.

Treatment

Best care is likely to be achieved by a multidisciplinary team including medical staff, specialist nurses, nutritionists and stoma-therapists with social and psychological support. Collaboration between gastroenterologist and surgeon is essential and should include patient sharing where appropriate, joint outpatient consultations for difficult cases, and early involvement of the surgeon in acute disease.

Medical treatment

Medical treatment involves bed rest and the correction of water and electrolyte depletion by intravenous infusion. Severely anaemic patients should be given blood. The patient should be encouraged to eat a high protein and calorie diet. Intravenous nutrition may be indicated in severely malnourished patients, as judged by a decrease in lean body mass and serum albumin.

Intravenous prednisolone (60 mg daily) and an H_2-receptor antagonist or a proton pump inhibitor to protect against upper gastrointestinal ulceration are given. Ciclosporin has been reported to induce remission in over 50% of patients unresponsive to steroids but early relapse may occur, resulting in the same clinical situation within a short period.

More recently, a number of studies have reported the use of biological agents for the treatment of acute ulcerative colitis. Infliximab (Remicade Centocor, Malvern, PA), a chimeric (75% mouse, 25% human) anti-tumour necrosis factor (TNF)-α monoclonal antibody, and Adalimumab (Humira, Abott), a humanized anti-TNF antibody, are currently used to modulate the proinflammatory processes central to the pathogenesis of IBD. In a study of 30 patients with active ulcerative colitis treated with infliximab between 2000 and 2006 at Oxford, 53% of patients came to colectomy at a median time of 140 days after their first infusion (range 4–607). Of those avoiding colectomy, only 17% sustained a steroid-free remission. The role of immunotherapy in the acute setting remains to be established in a number of randomized controlled trials.

Patients with severe acute colitis require admission to hospital. Initial treatment is medical but about 30% of patients will come to surgery. Surgery is absolutely indicated in cases with acute toxic dilatation or perforation.

Monitoring is essential to assess improvement or deterioration. The pulse rate, temperature and blood pressure are regularly recorded. The patient should be weighed on admission and twice weekly thereafter. Blood should be sent for haemoglobin, albumin and electrolyte estimations. A stool chart is essential. This should record every defecation with an assessment of volume and consistency of stool and the presence or absence of blood on each occasion. The abdomen should be examined regularly. Distension suggests the possible development of toxic megacolon. A plain radiograph will allow assessment of the colonic diameter. Abdominal tenderness and rigidity suggests local or general peritonitis. The presence of intramural gas on the plain abdominal radiograph is a sign of imminent perforation and is therefore an indication for immediate surgery.

Unresponsiveness to medical treatment in acute colitis

Failure to respond to medical treatment should be recognized early. The gastroenterologist and surgeon should confer at least daily to decide whether there has been improvement, stagnation or deterioration. Deterioration despite adequate medical treatment should be an indication for surgery. Stagnation over several days with no sign of improvement should also be an indication for operation.

Clinical indicators that surgery is likely to be necessary at the time of admission include a frequency of defecation of over 10 times per 24 hours with the passage of blood at every defecation attempt, low albumin, low haemoglobin and a fall in lean body mass of more than 10%. Previous acute attacks, poor general health and social circumstances affected by the disease should be taken into account. A significant history of chronic illness and social incapacity should sway the decision in favour of surgery.

Medical treatment of proctitis

Most patients are satisfactorily treated medically by a combination of steroids and 5-ASA preparations. The former is intended to induce a remission, the latter to maintain a remission once achieved. Both can be given as suppositories or as an enema, the choice depending on the proximal extent of disease. Steroid preparations such as budesonide have a lesser tendency for absorption. An oral 5-ASA preparation should also be prescribed from the beginning. Modern 5-ASA drugs (Asacol, Pentasa and Balsalazide) no longer contain sulphonamide, which was responsible for some of the side effects of salazopyrine. They are formulated to protect the aspirin from degradation before it arrives in the colon. Proctitis refractory to this treatment may respond to other preparations, including bismuth, nicotine and witchhazel. Rarely, patients with persisting severe symptoms may require surgery.

Megacolon, perforation and bleeding

Megacolon is an indication for surgery (Figure 30.33). If signs of peritonism are present, operation should not be delayed. When surgery is performed prior to perforation, the reported mortality is 2–8%. Perforation is a grave development with a mortality around 40%. It may be silent in a patient on large doses of steroids, and may become evident only by the presence of free gas on the plain abdominal radiograph. Its occurrence without megacolon is rare and should raise the possibility of Crohn's disease. Severe bleeding usually arises from ulceration in the rectum.

Medical treatment of chronic proctocolitis

Medical management includes anti-inflammatory, nutritional, symptomatic and psychological treatments.

Prednisolone is given in an initial dose of 40 mg, gradually reducing this as remission occurs over the next few weeks. Azathioprine can be tried in patients who do not respond to steroids or in those who are steroid dependent in the hope of avoiding long-term steroid treatment. Ciclosporin can be effective in inducing remission in patients suffering from acute

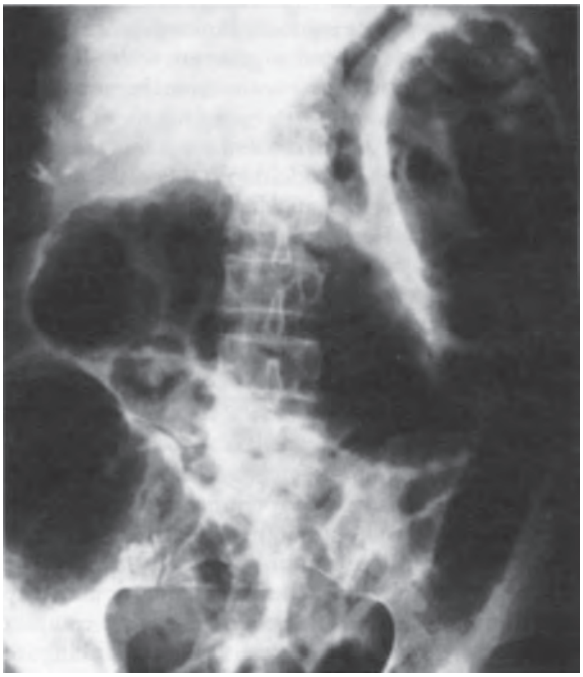


Figure 30.33 Plain radiograph appearances of toxic megacolon.

severe colitis but its role in chronic disease is not established. The nutritional state should be monitored. There is no specific diet that influences the activity of the disease but a high protein and calorie intake should be encouraged. Specific replacement treatment such as iron may be necessary. Antidiarrhoeal agents including codeine phosphate and loperamide are usually effective in reducing frequency and urgency. Maximal doses are 60 mg four times daily and 8 mg four times daily respectively. Lomotil (atropine and diphenoxylate) is occasionally effective where there has been a poor response to the others. Bone densitometry should be carried out where steroid medication has been prolonged.

Unresponsiveness to medical treatment in chronic colitis and surgical indications

Most patients requiring surgery have extensive disease. They are those most likely to have severe local symptoms and systemic illness. There is a greater risk of acute severe colitis and malignant transformation. Very occasionally, a patient with distal disease may require surgery, usually because of severe local symptoms. The indications for elective surgery include:

- unresponsiveness to medical treatment
- retardation of growth in the young
- malignant transformation.

Unresponsiveness to medical treatment

This indication can be further divided into various clinical categories, as follows.

Chronic disease

The patient continues to suffer from systemic and local symptoms despite adequate medical treatment. Chronic anaemia associated with general weakness, poor energy levels, amenorrhoea and extra-alimentary manifestations may leave the patient unable to lead a normal life. Patients may have experienced multiple hospitalizations, periods off work, disruption of family life and education and other social effects of chronic illness. Patients who have never experienced a complete remission from medical treatment are included in this group.

Steroid dependence

A response to steroids may be maintained only by continuing the therapy with relapse on withdrawal. If alternative medication such as immunosuppression is unsuccessful, then surgery is indicated unless there are particular reasons against.

Recurrent acute exacerbations

The decision for surgery will depend on the frequency and severity of attacks. Surgery during an acute attack will usually take the form of a colectomy with ileostomy. A decision taken during remission may allow an elective definitive procedure such as restorative proctocolectomy to be performed as the first stage procedure.

Severe symptoms

The patient may be systemically well but severely inconvenienced by frequency and urgency of defecation particularly if associated with urge incontinence.

Extra-alimentary manifestations

Not all symptoms will respond to removal of the large bowel; liver manifestations and sacroiliitis do not. However, the activity-related polyarthropathy does respond, as will some cases of pyoderma gangrenosum, although the latter may improve only slowly over several months.

Growth retardation

Ulcerative colitis, if extensive, has an inhibitory effect on growth and the development of secondary sexual characteristics. Steroid medication itself leads to early fusion of epiphyses, resulting in permanent stunting of growth. Patients in this category are usually under the care of a paediatric expert in the assessment of growth. However, within the years of puberty a delay in surgery may occur, partly because of the antipathy of the paediatrician and/or patient (or parents) to an ileostomy.

Malignant transformation

The presence of high- or low-grade dysplasia or an established invasive tumour is an indication for surgery. The surgical technique should be as though invasion had occurred since this can be determined only by examination of the resected specimen.

Surgical treatment

Acute colitis (emergency surgery)

Surgery has a major role in the management of acute colitis. The need for surgery is greatest during the first year after onset of the disease. The indications and their relative frequency are shown in Table 30.8 and Figure 30.34, respectively.

Colectomy with ileostomy and preservation of the rectum

This is the operation of choice for acute severe colitis. For all surgery for ulcerative colitis the reversed Trelendenberg position with the legs raised (Lloyd-Davies) should be used,

Table 30.8 Indications for emergency surgical intervention for ulcerative colitis 1976–1990

Main reason for surgery	No. of patients
Unresponsiveness to medical treatment	71
Toxic dilatation	23
Perforation	9
Bleeding	2
Other	1
Total	106

From Melville *et al.* *Gut* 1994;**35**:1076–80, with permission.

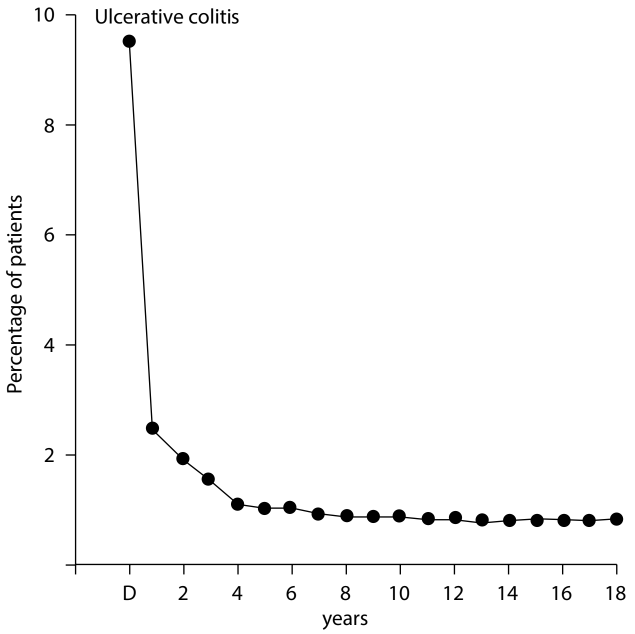


Figure 30.34 Colectomy rates in progressive years following the diagnosis (D) of ulcerative colitis.

thereby allowing access to the rectum. The bladder is routinely catheterised. It is helpful to insert a proctoscope before starting to drain the rectum and deflate the bowel.

Ileostomy trephine and incision

When an ileostomy forms part of the procedure, the trephine should be made before opening the abdomen. For open colectomy this should be midline. This can be limited to 7 cm or less in a thin patient. A paramedian incision is no longer appropriate. For laparoscopic colectomy a small incision is made to remove the specimen. It also allows a manual port if hand-assisted laparoscopy is used.

On opening the abdomen, care must be taken to avoid perforation. Where adhesions have formed between the colon and the parietes, dissection should be made within the latter.

Operative steps

These are discussed below.

- Mobilization of the right colon. The surgeon stands on the patient's left side.
- Division of the bowel. The bowel is divided at the ileocaecal junction before proceeding further. This allows control of the right colon manually to allow safe division of the ileocolic and right colic vessels, by avoiding tension.
- Division of vessels of the transverse colon.
- Mobilization of the left colon. The surgeon moves to the other side of the table to do this.
- Mobilization of the splenic flexure. The splenic flexure is often drawn down owing to shortening of the bowel due to the disease process and may be very easy to mobilize.
- Division of the sigmoid.

The level of division should allow sufficient length of distal bowel to be able to exteriorize it through the anterior abdominal wall whether a mucous fistula is formed or not. Division at the level of the peritoneal reflection leaves a distal stump that is too short to be exteriorized in the uncommon event of breakdown of the distal suture line and also makes identification of the rectum at a subsequent operation difficult. A long rectosigmoid stump should therefore be aimed for unless it is necessary to remove the rectum in the case of severe bleeding (Figure 30.35).

Mucous fistula formation

Although leakage of the distal closure suture line is uncommon (<5%), the creation of a mucous fistula should be considered in patients in poor condition with malnutrition and where the bowel wall is too diseased safely to take sutures. The ileostomy and mucous fistula should be brought out sufficiently far apart to allow a stoma appliance to be placed over each without interference. Placing the mucous fistula in the contralateral iliac fossa achieves this. Some surgeons prefer to close the stump and leave it in the subcutaneous fat deep to the abdominal wound.

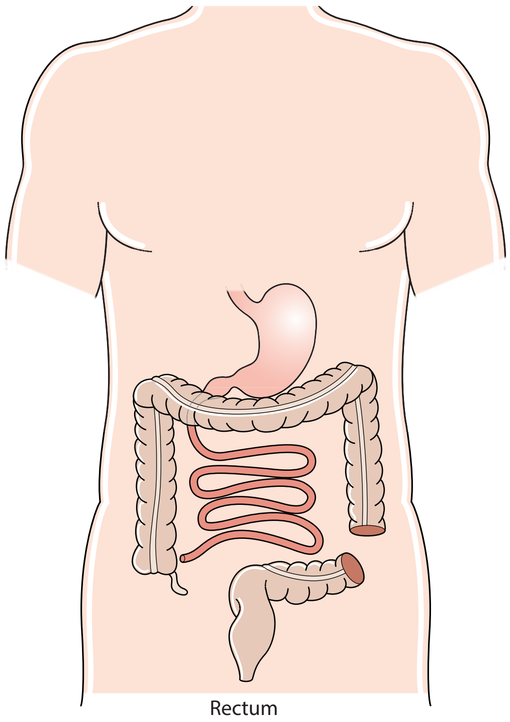


Figure 30.35 Emergency colectomy: level of division of the sigmoid allowing adequate mobility for exteriorization.

Postoperative outcome

In the immediate postoperative period, it is advisable to drain the rectum (particularly if the stump is closed) by daily insertion of a proctoscope. Complications include intestinal obstruction and sepsis. Obstruction usually resolves spontaneously. Sepsis may be due to an intra-abdominal collection or leakage from a closed rectosigmoid stump. This will usually require reoperation and exteriorization of the distal bowel if possible. Rarely, rectal excision preserving the anal canal will be necessary. Occasionally, persisting inflammation in the rectal stump may be so severe despite local treatment that rectal excision is required.

Recovery usually occurs within a few weeks to a couple of months, during which steroids and other medication are gradually withdrawn. Health and self-confidence are restored allowing return to work or education. The timing of any subsequent operation is determined by the patient but should not be before 3 months.

Mortality ranges from <1% to 5% from specialist centres in the UK, Denmark and the USA. Recent linkage analysis of 8245 patients with ulcerative colitis having a first hospital admission for ulcerative colitis of more than 4 days in England between 1998 and 2000 indicate that the mortality up to 2003 (3 years) is lower after an elective (2.2%) or emergency procedure (3.2%) than if no operation is carried out (7.4%). If substantiated these data suggest that the threshold for surgery may be too high. There is evidence that mortality is higher in smaller, low - volume hospitals.

Chronic colitis (elective surgery)

There are four surgical options:

- colectomy with ileostomy and preservation of the rectum
- conventional proctocolectomy and permanent ileostomy
- colectomy and ileorectal anastomosis
- restorative proctocolectomy and ileal reservoir.

All these operations involve total removal of the colon (Figure 30.36). There is no place for partial colectomy, even in cases with a normal right colon, where experience has shown a high frequency of recurrence in the remaining colon. In patients having a permanent ileostomy, a Kock pouch (see below) may be possible in certain circumstances.

Colectomy with ileorectal anastomosis should be considered only where the rectum is minimally inflamed and distensible and where there is no evidence of dysplasia anywhere in the large bowel. Most patients do not fulfil these criteria, leaving the alternative of conventional or restorative proctocolectomy. The latter is excluded, however, if the anal sphincter is inadequate or the anus is diseased. Much will depend on the wishes of the patient. Here, it is essential for the clinician to give detailed information on the morbidity, function, late complications and likely duration of treatment of the various options. Input from the stomatherapist and patient-support groups should be obtained. The introduction of restorative proctocolectomy has resulted in this operation being the most commonly used with over 80% of patients requiring elective surgery.

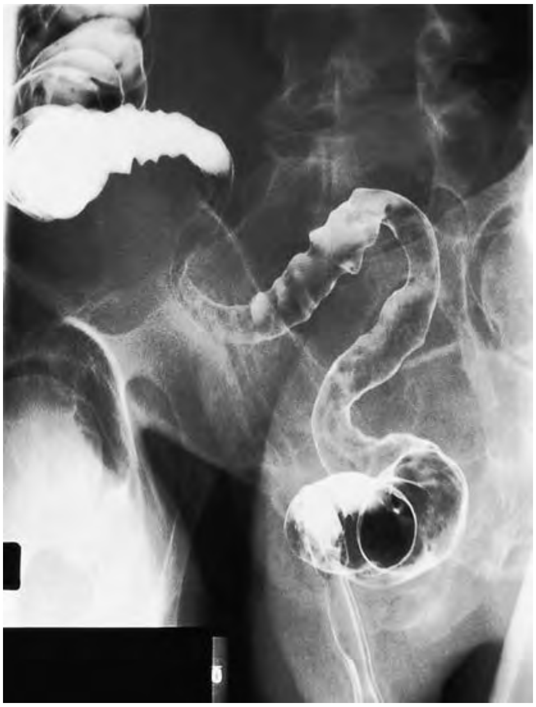


Figure 30.36 Barium enema appearances of longstanding ulcerative colitis.

Colectomy with ileostomy and preservation of the rectum

This operation also has a place in elective treatment. For patients too unfit for a restorative proctocolectomy it may be preferable to carry out an initial colectomy. Besides the advantages of rapid return to health (see above), the experience of an end-ileostomy is useful to the patient when considering further surgery. Where the diagnosis of ulcerative colitis is in doubt, a colectomy gives the histopathologist more material to make a diagnosis. Dysplasia is extremely rare within 10 years of the onset of colitis and there is, therefore, no hurry for reoperation in most cases. All of the options discussed below can then be considered.

Conventional proctocolectomy with permanent ileostomy

The introduction of the everted ileostomy established this operation as the standard procedure for ulcerative colitis until the description of restorative proctocolectomy. Ulcerative colitis is cured, with the only disadvantage being a permanent ileostomy.

Indications

The indications include rectum and anus not suitable for restorative procedure, i.e.:

- inadequate sphincter
- cancer in the low rectum
- patient preference.

The patient should appreciate the possibility of ileostomy-related complications requiring further surgery and delayed healing of the perineal wound. These are offset by those of pelvic sepsis and long-term developments such as pouchitis after restorative proctocolectomy.

Operative steps

Creating the ileostomy trephine before the midline incision and colectomy have already been discussed. Where a carcinoma or dysplasia is present a conventional anatomical dissection total mesorectal excision (TME) should be carried out. The majority of patients do not have neoplastic transformation and, in these, the rectum may be removed by close (perimuscular) dissection to minimize the incidence of pelvic nerve damage, which might cause urinary or sexual dysfunction although there was no statistically significant difference in a non-randomized comparison of perimuscular and conventional dissection. Whatever technique is used, dissection should be kept behind Denonvilliers' fascia anteriorly and close to the rectal wall laterally where the autonomic pelvic nerves are mostly at risk. In the absence of carcinoma in the low anorectal region, the anal canal should be removed using the intersphincteric technique. Good ileostomy technique includes preparation of the mesentery and its fixation to the peritoneum of the anterior abdominal wall. The terminal 5 cm of ileum are perfused by the marginal vessel after division of the ileocolic artery and vein. The stoma itself is completed by mucocutaneous interrupted sutures. It will evert spontaneously; a projection of 2.5 cm is ideal.

Postoperative outcome

Obstruction may occasionally require reoperation but usually settles spontaneously. Rarely it is due to herniation of the small bowel into the space lateral to the ileostomy. A haematoma in the cavity left by the rectum may become infected. Drainage with resolution of the acute problem may be followed by a perineal sinus. This may require subsequent surgery with curettage and rarely in refractory cases the need for a perineal myocutaneous rectus abdominus flap. Delayed healing of the perineal wound at 6 months occurs in 10–20% of cases

Ileostomy complications are common. Stricture formation, prolapse and retraction may cause difficulty in maintaining a water - tight appliance, as may parastomal herniation. Corrective surgery is often required, with a reported cumulative ileostomy revision rate of around 25% at 5 years. It is usually possible to carry out a local revision, but in some cases, particularly with herniation, resiting is required. This is a major undertaking.

The continent ileostomy

Kock developed this operation (Figure 30.37) initially for bladder replacement and then adapted it to create a continent abdominal intestinal stoma. The procedure still has a place in patients in whom the anal sphincter has been removed. A reservoir is constructed from 30 cm of the terminal ileum and the most distal 15 cm of small bowel are invaginated into the reservoir to form a nipple valve. This maintains continence and the reservoir is emptied several times per day by catherization of the abdominal stoma using a wide - bore tube.

Early complications include leakage from the reservoir causing peritonitis or fistulation. The most important late complication is subluxation of the nipple valve. It is suggested by the onset of incontinence of the stoma and difficulty in inserting the catheter. Contrast radiology may show partial or complete prolapse of the valve. Valve slippage occurs in 17% to over 40% of cases, and is the most common cause of failure. Reported fistula rates range from 10% to 26%. Of 330 patients treated at a single institution between 1974 and 2001, pouch survival at 10 and 20 years was 87% and 77%, respectively. Patients with indeterminate colitis or Crohn's disease were 4.5 times more likely to lose the pouch than those without, although in one series the cumulative 10 year recurrence rate in 49 patients with Crohn's disease was 48%, with excision of the pouch required in only eight.

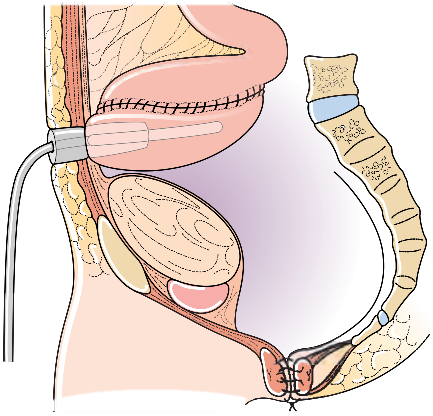


Figure 30.37 The Kock continent colostomy.

Long-term continence rates of over 90% have been reported. The cumulative 4 year incidence of pouchitis is around 40%. Dysplasia has been reported in three of 40 patients followed for a median of 30 years. One case of carcinoma in the pouch has been reported.

Continent ileostomy may be indicated in motivated patients in whom restorative proctocolectomy is not an option and in patients who already have a permanent ileostomy and desire an improved quality of life. It has also been carried out after failure of restorative proctocolectomy.

Colectomy with ileorectal anastomosis

Indications

Colectomy with ileorectal anastomosis is a well-tolerated one-stage procedure. Before the introduction of restorative proctocolectomy it was the only procedure which avoided an ileostomy being used in 10% to over 80% of patients according to the preference of the surgeon. Since then, its incidence has fallen to below 10%. The indications are as follows:

- a non- or mildly inflamed rectum with good compliance
- absence of dysplasia anywhere in the large bowel
- adequate anal sphincter
- availability of the patient for follow-up
- presence of disseminated colonic carcinoma with relative rectal sparing.

The patient must also be prepared to be followed by annual rectoscopy with biopsy owing to the risk of malignancy in the rectal stump which is about 5% at 20 years. If this is not possible then the operation should not be advised.

Operative steps

The technique is identical to colectomy with ileostomy up to the point of removal of the specimen. The upper rectum is very accessible for manual or stapled anastomosis (Figure 30.38).

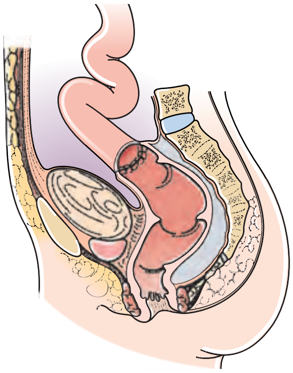


Figure 30.38 Colectomy with ileorectal anastomosis.

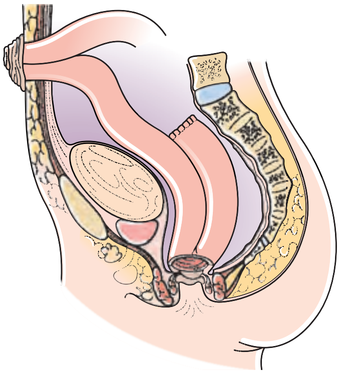


Figure 30.39 Ileal J-pouch with proximal defunctioning ileostomy.

Postoperative outcome

The operation is a compromise since there is potentially active disease in the rectum. This can lead to failure because of persisting inflammation causing poor function or the development of malignancy. The results demonstrate a low mortality and morbidity but considerable differences in failure requiring removal of the rectum. There were 12 deaths among 22 cancers that developed during follow-up in a series of 384 patients.

Patients who fail are candidates for rectal excision with permanent ileostomy or restorative proctectomy, provided that any cancer in the rectum can be adequately cleared without having to sacrifice the anal canal and provided there is no dissemination.

Restorative proctocolectomy with ileal reservoir

The strategy of complete removal of the large bowel with sphincter preservation was first described fully by Ravitch and Sabiston. The 'straight' ileoanal anastomosis was adopted by a few surgeons at the time and the functional results were subsequently reviewed by Valiènte and Bacon. Function was often poor due largely to urgency and frequency.

The introduction of the continent ileostomy showed that a small bowel reservoir could function in humans and this led Parks to combine this with his own endoanal anastomotic technique to create the ileoanal ileal reservoir procedure (Figure 30.39). Capacitance of the neorectum is inversely related to frequency, irrespective of whether the neorectum is constructed using straight ileum, or an ileal or colonic reservoir

Indications

The only reason for the operation is to avoid a permanent ileostomy. A conventional proctocolectomy gives excellent results except for this. Where there is no medical objection, the choice lies between a restorative or conventional proctocolectomy and is almost entirely the patient's to make. This is possible only if the disadvantages are fully discussed. These include failure and complication rates, total treatment time, the possibility of pouchitis occurring and the likely functional outcome. A pouch support nurse, stomatherapist and patient support group can offer valuable advice but in the end the patient must decide.

The indications include:

- ulcerative colitis
- familial adenomatous polyposis (FAP)
- non-acute severe colitis; any severely ill patient should have an initial colectomy
- absence of low rectal cancer; where a cancer is present, it must be possible to achieve locoregional clearance, an assessment similar to that made when considering neoadjuvant chemoradiotherapy or anterior resection or total rectal excision for 'ordinary' rectal carcinoma. The operation is not indicated in patients with disseminated disease
- adequate anal sphincter; manometry should be performed where there is clinical uncertainty of sphincter function.

Surgical controversies

Age

Failure, complication rates and function are similar in paediatric patients to overall series. General health and quality of life is similar to that of healthy children in those patients with a functioning pouch. There is some evidence that incontinence, usually minor, is more common in patients over 45 years. There is, however, no absolute contraindication in older patients and the decision should depend on assessment of the individual patient, particularly regarding sphincter function.

Female fertility

Female infertility is increased by two to three times following restorative proctocolectomy. This applies to both ulcerative colitis and FAP but more so in the former. There is also a higher incidence of fertility treatment among patients after restorative proctocolectomy than in the normal population. Fertility is not affected by colectomy. The pelvic dissection is one factor responsible as well as the advancing age of the patient. Fertility is of medicolegal importance when counselling females of child-bearing age. The patient may decide to have a restorative proctocolectomy accepting the risk of reduced fertility or she may decide to have a colectomy with ileostomy allowing recovery from the disease while preserving fertility. A restorative proctectomy can then be considered at a later date convenient to the patient.

Crohn's disease

With the exception of one group, patients with Crohn's disease experience a higher failure rate than the general pouch population. In one series failure rates for ulcerative colitis and Crohn's disease were 10% and 50%, respectively. Over a 20-year period of follow-up failure occurred in 55%. Failure in Crohn's disease is significantly greater than in indeterminate colitis.

Table 30.9 Outcome of restorative proctocolectomy in Crohn's disease

Study	N	Failed (%)
Galandiuk SS *et al.* (1990)	16	4
Deutsch AA *et al.* (1991)	9	4
Hyman NH *et al.* (1991)	25	7
Grobler SP *et al.* (1993)	20	8
Panis Y *et al.* (1996)	31	2
Regimbeau *et al.* (2001)	41	3
Tulchinsky H *et al.* (2003)	13	6
Hahnloser DPJ *et al.* (2007)	44	20

Thus, as a general principle, Crohn's disease is a contraindication to restorative proctocolectomy (Table 30.9).

Indeterminate colitis

Indeterminate colitis without radiological or clinical evidence of Crohn's disease tends to behave like ulcerative colitis. Failure rates in large series followed for 10 or more years are around 10% and at 20 years there was no significant difference in the failure rates among patients with ulcerative or indeterminate colitis (6% and 12%). Complications and function also appear to be similar to patients having the operation for ulcerative colitis. Exclusion of small bowel and anal disease is obligatory.

Previous anal pathology

A history of anal disease may suggest Crohn's disease. The presence of an anal lesion increases the risk of anastomotic leakage, pouch – vaginal and perineal fistula and subsequent failure.

Sclerosing cholangitis

Patients with sclerosing cholangitis have double the incidence of pouchitis after restorative proctocolectomy. They are also at increased risk of dysplasia developing in the pouch, although this is small. While not a contraindication, the patient should be carefully counselled. Liver function tests should be performed routinely preoperatively.

Operative steps

The steps in the operation include:

1. mobilization of the colon and rectum: this is identical to conventional proctocolectomy
2. division of the gut tube
3. the rectum is mobilized to the anorectal junction. If a stapled ileoanal anastomosis is intended, a transverse stapler is applied at this level. Where a manual anastomosis is to be carried out, the bowel is divided, leaving an open anal stump
4. mobilization of the mesentery
5. having removed the surgical specimen, the small bowel mesentery is fully mobilized. It is useful to perform a trial descent to the anal canal of the point on the ileum selected for the ileoanal anastomosis. If further mobilization is necessary, this may require the division of selected vessels but care must be taken to avoid ischaemia
6. ileal reservoir formation
7. anastomosis
8. defunctioning ileostomy.

The two-loop (J) reservoir is easy to make by hand or by stapling. The original three – loop (S) reservoir of Parks often led to evacuation difficulty due to the short segment of ileum distal to the pouch. The four-loop (W) reservoir achieves a capacitance generally greater than that of the J reservoir and similarly has no distal segment (Figure 30.40). In general, the larger the reservoir, the lower the stool frequency. An operative image of the J reservoir is shown in Figure 30.41.

Anastomosis

Manual

A mucosectomy is performed via an endoanal approach. Using suitable retraction (e.g. the Lone Star) the mucosa is removed by sharp scissor dissection after elevation by the submucosal injection of saline with adrenaline (1/300 000). The reservoir is then brought through the anal canal and sutures (12 sutures, one for each hour of the clock) are placed after removal of the retractor to minimize tension.

Stapled

A transverse stapler is applied at the level of the anorectal junction and the bowel is then divided. The anvil of a circular stapling instrument is fixed in the reservoir by a pursestring suture and the anastomosis completed on firing the instrument after insertion *per anum*.

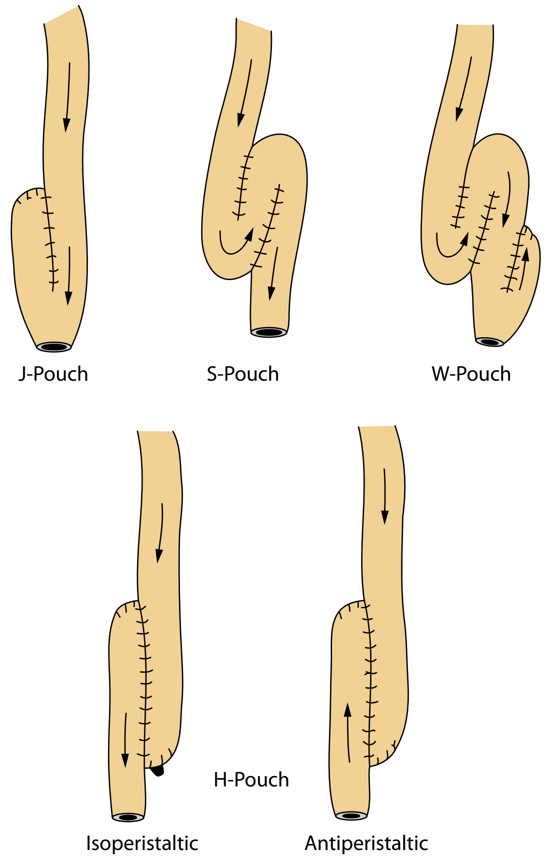


Figure 30.40 Diagrammatic representation of the various forms of pouches used after colectomy and mucosal proctectomy. The J pouch is the most extensively used since it is the simplest to create, is attended with the lowest incidence of complications and functions best in the long term.

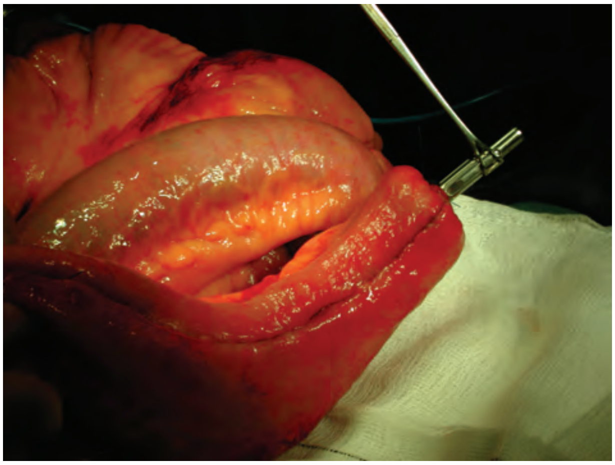


Figure 30.41 Pouch anastomosis using the J-pouch configuration

The manual technique allows a precise level for the anastomosis and avoids the possibility of leaving rectal mucosa behind, while stapling may sometimes result in an anastomosis to the rectum leaving a length of distal proctitis. This may cause continuing bleeding, discomfort, urgency and evacuation difficulty. 'Strip proctitis' or 'cuffitis' has been reported in around 10–15% of cases having a stapled anastomosis with a small cumulative risk of dysplasia of 3% over a mean follow-up of 16 months in one study. A few cases of carcinoma distal to the ileoanal anastomosis have been reported but these have occurred in patients who had either dysplasia or invasive cancer in the original operative specimen.

A stapled anastomosis is easier and quicker to perform and it may cause less trauma to the anal sphincter and is preferred for patients in whom there may be tension in the mesentery on bringing the reservoir down to the anal level. A recent meta-analysis of 21 comparative studies has demonstrated no significant differences in the incidence of postoperative complications between the two types of anastomosis. The incidence of nocturnal seepage and pad usage favoured the stapled anastomosis but persisting symptoms due to inflammation or dysplasia favoured the manual technique. The surgeon should be capable of using either method (Figure 30.42).

Most surgeons include a defunctioning ileostomy routinely. The ileostomy can cause morbidity, however, in both its formation and its closure and this may account for 20% of complications. The operation has been carried out without ileostomy with excellent results. However, in a meta-analysis the rate of anastomotic leakage and the subsequent development of an anal lesion was double in patients not given a defunctioning ileostomy.

Postoperative outcomes

Failure is defined as the need for excision of the reservoir or indefinite diversion. The learning curve in ileal pouch surgery is related to failure which improved following an initial training period of 23 cases in one study. Adequate results for manual anastomosis followed an initial period of 31 procedures.

Reports on failure in the early years after closure of the ileostomy gave rates ranging from around 5% to 10%. Over the longer term failure continues in a linear manner with rates approaching 15% at 15 years although in another large study a failure rate of 3.7% at 5 years had risen to only 7.9% at 20 years (Table 30.10).

The reasons for failure include pelvic sepsis (50%), poor function (30%) and pouchitis (10%). Pelvic sepsis in the early postoperative period confers a fivefold increase in the chance of subsequent failure. In a series of 1965 patients treated in a single centre between 1983 and 2001, four preoperative and four postoperative factors were found to be associated with ileal pouch failure. Using a multifactorial survival analysis, each

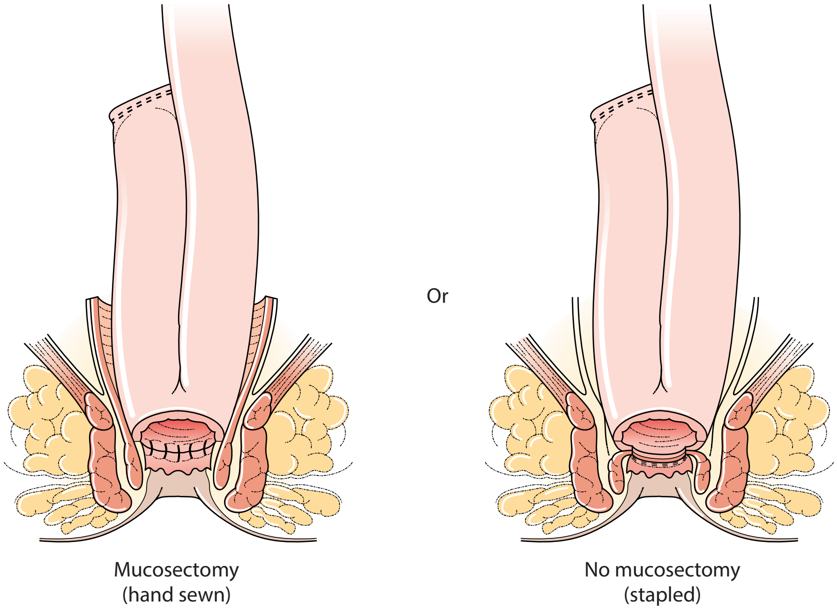


Figure 30.42 Comparison of the hand-sewn anastomosis with mucosectomy (a) and stapled anastomosis without mucosectomy (b).

Table 30.10 Failure rate after restorative proctocolectomy

Study	***N***	**Follow-up (months*)**	**Pouch excision**	**Indefinite diversion**	**Overall failure (%)**
Gemlo BT *et al.* (1992)	253	>12	–	–	9.9
Foley EF *et al.* (1995)	460	–	7	9	3.5
MacRae HM *et al.* (1997)	551	>30	49	9	10.5
Korsgen S *et al.* (1997)	180	>24	23	8	17.2
Meagher AP *et al.* (1998)	1310	24–180 (77)	84	50	10
Tulchinsky H *et al.* (2003)	634	36–288 (85)	41	20	9.7
Fazio V *et al.* (2003)	1975	1–228 (49)	38	39	4.1

* Numbers in parentheses indicate mean values.
† Patients with ulcerative colitis only.

factor was assigned appropriate weights to form a scoring system for quantifying the risk of pouch failure in individual patients (Table 30.11).

Complications

Morbidity ranges from 20% to 50%. Many complications resolve spontaneously but some may require active intervention.

Pelvic sepsis

Pelvic sepsis owing to breakdown of the ileoanal anastomosis or an infected haematoma, or both has been reported in 5–20% of patients. A pyrexia develops within a few days and digital examination *per rectum* may reveal an anastomotic defect or an extraluminal swelling (usually posterior) indicating the presence of a collection. The passage of fresh blood is highly suggestive of sepsis. A CT scan and examination under anaesthesia should be performed and any collection drained into the lumen.

Stricture

Stricture of the anastomosis requiring active intervention (either dilatation or a more major procedure) has been reported in 15% to over 30%.

Intestinal obstruction

This occurs in 5–20% of patients and surgery may be necessary, although most cases resolve spontaneously.

Pouch–vaginal fistula

Pouch–vaginal fistula (the male equivalent is pouch–perineal fistula) occurs in 5–10% of (female) patients and is an important reason for late failure. It can occur months or years after closure of the ileostomy and is associated with an anastomotic complication in most cases. Patients with indeterminate colitis or Crohn's disease have a significantly higher incidence ulcerative colitis patients (hazard ratio 1.4, 2.2 respectively). The presence of perianal abscess or fistula *in ano* preoperatively is associated with 3.7- and sixfold increases in the risk of developing pouch–vaginal fistula.

Function

Frequency ranges from a median of four to seven defecations per 24 hours, but 20–30% of patients have a frequency of eight or more. This is usually regarded by the patient as acceptable, probably because urgency is present in less than 5% of patients. Nocturnal defecation is probably the most sensitive indicator of function. Frequency varies spontaneously and is also influenced by diet. Continence rates also vary, but faecal incontinence is rare (5%). The need for antidiarrhoeal medication ranges from 20% to 50%.

Table 30.11 (a) Ileal pouch failure model and (b) conversion chart for the prediction of ileal pouch failure following restorative proctocolectomy

(a)	**Points**
Preoperative risk factors	
Diagnosis	
Familial adenomatous polyposis	0
Ulcerative colitis or indeterminate colitis	1
Crohn's disease	1.5
Patient comorbidity	
No comorbid conditions	0
One comorbid condition	0.5
Two or more comorbid conditions	1.0
Prior anal pathology	
No prior anal pathology	0
Prior anal pathology	1
Anal sphincter manometry	
Normal manometry	0
Abnormal manometry	1
Postoperative risk factors	
Anastomotic separation	
No anastomotic separation	0
Anastomotic separation	1
Anastomotic stricture	
No stricture or asymptomatic stricture	0
Symptomatic stricture	1
Pelvic sepsis	
No sepsis	0
One episode of pelvic sepsis	1
Two or more episodes of pelvic sepsis	2
Fistula formation	
No fistula	0
Pouch–perineal fistula	1
Pouch–vaginal fistula	2

(b) Score	Follow-up (years)		
	1	5	10
0	0.1%	0.4%	0.8%
1	0.3%	1.1%	2.0%
2	0.8%	2.9%	5.0%
3	2.0%	7.2%	12.3%
4	5.0%	17.4%	28.5%
5	12.4%	38.7%	57.7%
6	28.7%	71.5%	89.0%

(a) From Fazio *et al.* *Ann Surg* 2003; **238**:605–14, with permission.

There is a tendency for function to improve with time with regards to bowel frequency; however, seepage and minor incontinence tend to worsen with time. Based on the results of the Association of Coloproctology of Great Britain and Ireland (ACPGBI) Ileal Pouch registry of over 2500 pouch surgeries, the long-term frequency of defecation remained stable at a median of around six times per 24 hours and one per night when followed over 20 years. Minor incontinence (seepage) has increased from 4% to nearly 20% during the same period. Similar results for minor incontinence (3.9% at 1 year, 21% at 20 years) with a sight increase in 24 hour frequency (7.2 at 1 year, 8.4 at 20 years) have been reported by others. Function is therefore well maintained over time with a small decline in continence. See Table 30.12 for the long-term functional outcomes following restorative proctocolectomy based on the data from the ACPGBI Ileal Pouch Registry.

Long-term outcomes

General

Deficiencies of iron and vitamin B_{12} occur in less than 10% over a follow-up of 2–3 years. These may be associated with anaemia. There is a rise in the concentration of bacteria in the pouch by over a million-fold from the concentration of 10^4–10^6 colony-forming units per gram of faeces found in the normal terminal ileum.

Pouchitis

The small bowel mucosa develops a degree of villous atrophy in almost all cases (including those with FAP). The endoscopic appearances of a non-inflamed pouch are shown in Figure 30.43. In some patients, however, acute inflammation occurs and when this is associated with frequency, urgency, liquid stool and extra-alimentary manifestations the condition is known as pouchitis (Figure 30.43). There is evidence that pouchitis occurs early after closure of the ileostomy. It is rare in FAP and is more common in non-smokers and in patients with sclerosing cholangitis.

The cumulative incidence over 5 years approaches 50% but the prevalence of unremitting chronic pouchitis is only about 5%. The diagnosis depends on clinical, endoscopic and histopathological features, the last being essential and must show acute inflammation. Grading systems to assess severity have been described. A bacterial cause is likely to be partly responsible since pouchitis responds to antibacterials and probiotics. Antibacterial drugs, including metronidazole, ciprofloxacin and augmentin, can induce a response in over 80% of patients. Maintenance of remission by daily administration of the probiotic VSL3 has been shown to be effective in 85% of patients compared with controls (0%) when taken over a 9 month period. Withdrawal of treatment is followed by recurrence. Long-term metronidazole should be avoided due to the risk of peripheral neuropathy. An algorithm for the treatment of pouchitis is shown in Figure 30.44.

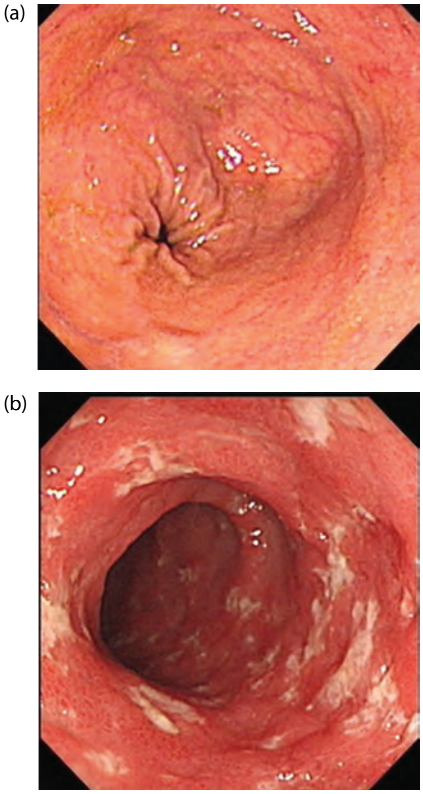


Figure 30.43 Normal pouch appearances (a) and pouchitis on endoscopy (b).

Table 30.12 Long-term functional outcomes of pouch formation

Functional outcome	1 year	5 years	10 years	15 years	20 years
Median stool frequency					
Per 24 hours	5	5	6	5	5
Night	0	0	1	0	1
Seepage – daytime	3.9%	3.8%	6.6%	7.2%	20.5%
Seepage – night	8.0%	8.2%	10.6%	10.1%	15.4%
Pad use – daytime	2.9%	2.5%	5.9%	5.8%	12.8%
Pad use - night	6.1%	5.0%	7.2%	8.7%	17.9%
Faecal urgency	5.1%	5.9%	9.4%	9.1%	2.9%
Antidiarrhoeal medication	38.9%	34.0%	37.2%	25.4%	38.5%

Removal of the reservoir for pouchitis accounts for only 10% of failures. Defunctioning does not affect the degree of inflammation, and excision with construction of a new reservoir is followed by pouchitis.

Neoplastic transformation

Dysplasia in the pouch mucosa is rare. To date over 20 case reports of carcinoma in the pouch or the distal anorectal segment have been described. Almost all of these had either dysplasia or a carcinoma in the original operative specimen and in none did the carcinoma appear within less than 10 years from the diagnosis of ulcerative colitis. These patients should be selected for surveillance by endoscopy with multiple biopsies at yearly intervals beyond 10 years from the onset of colitis. A suggested surveillance protocol is shown in Figure 30.45.

Salvage surgery

Removal of the pouch has a significant morbidity, with readmissions and delayed healing of the perineal wound in 40%

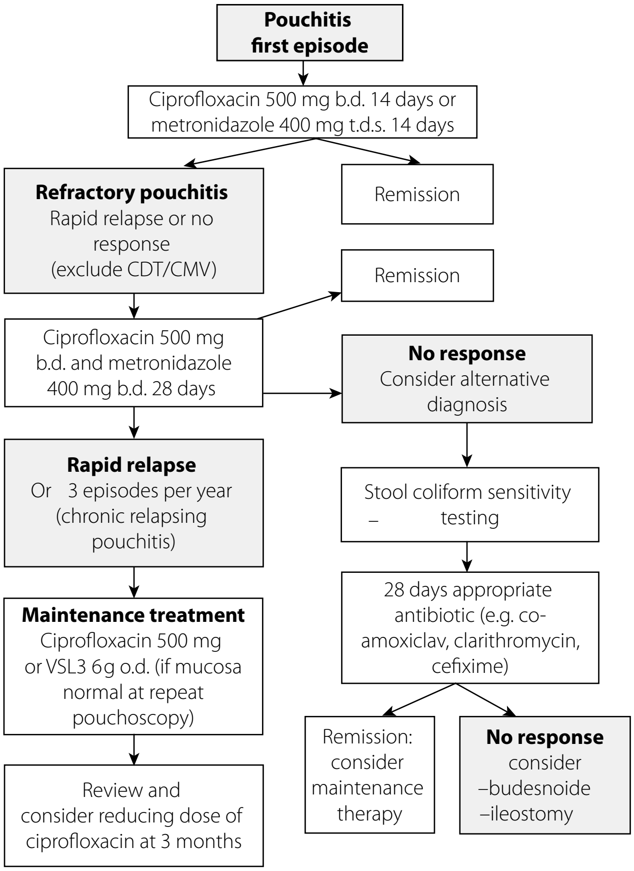

Pouchitis first episode
Ciprofloxacin 500 mg b.d. 14 days or metronidazole 400 mg t.d.s. 14 days
Refractory pouchitis
Rapid relapse or no response (exclude CDT/CMV)
Remission
Remission
Ciprofloxacin 500 mg b.d. and metronidazole 400 mg b.d. 28 days
No response
Consider alternative diagnosis
Rapid relapse
Or 3 episodes per year (chronic relapsing pouchitis)
Stool coliform sensitivity – testing
Maintenance treatment
Ciprofloxacin 500 mg or VSL3 6g o.d. (if mucosa normal at repeat pouchoscopy)
28 days appropriate antibiotic (e.g. co-amoxiclav, clarithromycin, cefixime)
Review and consider reducing dose of ciprofloxacin at 3 months
Remission: consider maintenance therapy
No response
consider –budesnoide –ileostomy

Figure 30.44 Algorithm for the treatment of pouchitis. (Adapted from McLaughlin SD *et al*. *Aliment Pharmacol Ther* 2008; **27**: 895–909.)

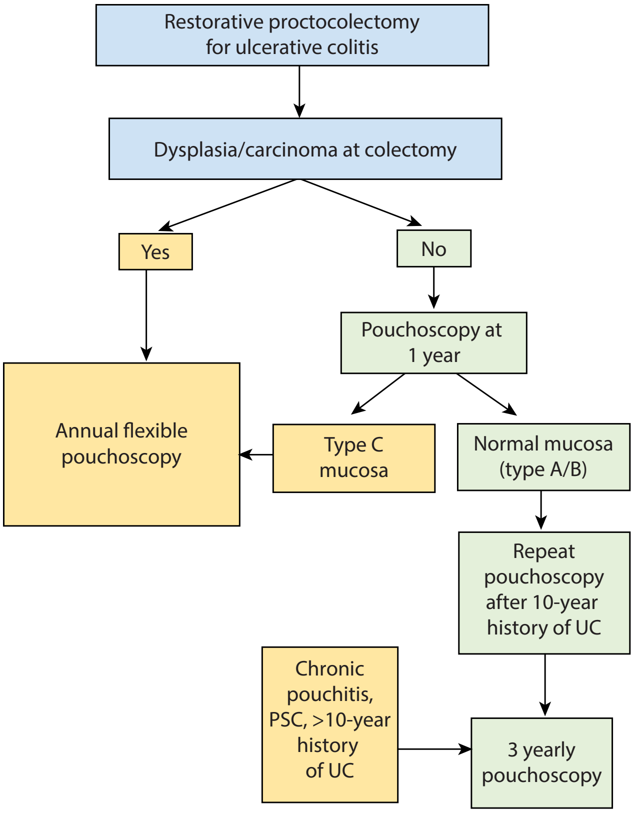

Restorative proctocolectomy for ulcerative colitis
Dysplasia/carcinoma at colectomy
Yes
No
Pouchoscopy at 1 year
Annual flexible pouchoscopy
Type C mucosa
Normal mucosa (type A/B)
Repeat pouchoscopy after 10-year history of UC
Chronic pouchitis, PSC, >10-year history of UC
3 yearly pouchoscopy

Figure 30.45 Suggested surveillance protocol algorithm for postoperative ulcerative colitis. PS C, UC, ulcerative colitis. Adapted from McLaughlin SD *et al*. (*Aliment Pharmacol Ther* 2008; **27**:895–907.)

of patients. Thus, where failure is threatened by sepsis or poor function, it may be in the patient's interest to consider a salvage procedure which may be less traumatic and offers a chance of retaining satisfactory anal function.

Such patients often have poor function and should be investigated by contrast radiology, including evacuation pouchography, anal manometry, pouch volumetry, MRI and histopathological examination of biopsy material. Action will depend on the diagnosis (Figure 30.46).

Success rates for abdominal salvage for pelvic sepsis range from <30% to 80%. Pouch–vaginal fistula should be treated by defunctioning followed by an attempt at repair. If low a local approach via the endoanal or transvaginal route will be necessary and success rates of around 60% of patients have been reported, although late recurrence can occur. If high (usually associated with a stapled ileoanal anastomosis), abdominal advancement of the anastomosis with mucosectomy has achieved success in over 70%. Pouch–perineal fistula in males is difficult to close but management by a long-term seton can be satisfactory.

Abdominal revision for mechanical outflow obstruction due to anastomotic stricture or a retained rectal stump is successful in 70–90%. This is significantly higher than the rates achieved when sepsis is the indication for salvage (Figure 30.47). Augmentation of a small volume reservoir may improve function.

The key points in the management of ulcerative colitis are listed below.

- The prevalence of ulcerative colitis is about 160 per 100 000 population compared with about 50 per 100 000 for Crohn's disease.
- At presentation, 50% of cases have disease confined to the rectum; in 30% this extends to the left colon, and in a further 20% disease extends beyond the splenic flexure.
- About 25% of patients with ulcerative colitis will develop at least one extra-alimentary manifestation during the course of the illness such as arthropathy, primary sclerosing cholangitis, erythema nodosum or uveitis.
- The incidence of cancer in patients is estimated to be <1% within 10 years of onset of ulcerative colitis, 10–15% in the second decade and over 20% in the third.
- Histopathological examination of a mucosal biopsy is the basis for the diagnosis of ulcerative colitis.
- The initial treatment of acute severe colitis is medical but about 30% of patients will require surgery. Surgery is absolutely indicated in cases with acute toxic dilatation or perforation.
- The operation of choice for acute severe colitis is subtotal colectomy with ileostomy and preservation of the rectal stump.
- Indications for elective surgery for chronic ulcerative colitis include failure of medical treatment, growth retardation in the young and neoplastic transformation.
- Restorative proctocolectomy is the procedure of choice for the majority of patients with chronic ulcerative colitis.

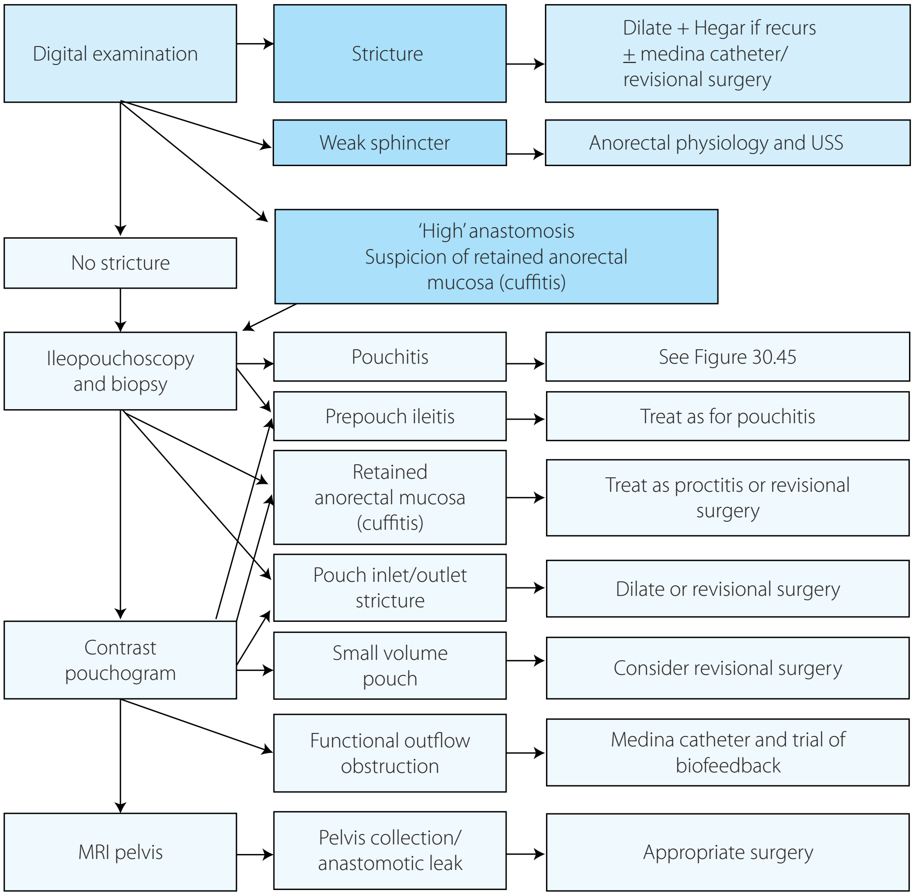

Digital examination
Stricture
Dilate + Hegar if recurs ± medina catheter/ revisional surgery
Weak sphincter
Anorectal physiology and USS
No stricture
'High' anastomosis Suspicion of retained anorectal mucosa (cuffitis)
Ileopouchoscopy and biopsy
Pouchitis
See Figure 30.45
Prepouch ileitis
Treat as for pouchitis
Retained anorectal mucosa (cuffitis)
Treat as proctitis or revisional surgery
Pouch inlet/outlet stricture
Dilate or revisional surgery
Contrast pouchogram
Small volume pouch
Consider revisional surgery
Functional outflow obstruction
Medina catheter and trial of biofeedback
MRI pelvis
Pelvis collection/ anastomotic leak
Appropriate surgery

Figure 30.46 Suggested algorithm for the investigation of pouch dysfunction. USS, ultrasound scan. (Adapted from McLaughlin SD *et al. Aliment Pharmacol Ther* 2008; **27**:895–909.)

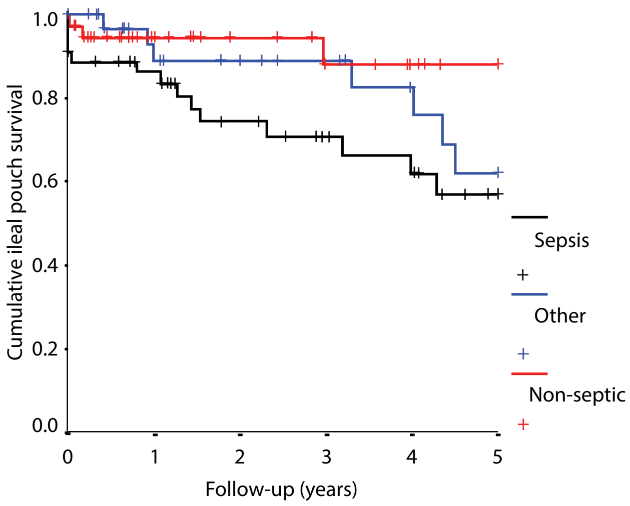


Figure 30.47 Anal function up to 5 years following abdominal salvage surgery for septic and non-septic indications

- Contraindications to restorative proctocolectomy include Crohn's disease, indeterminate colitis favouring Crohn's disease, weak sphincter and patient choice.
- The choice of ileal pouch reservoir (J vs W) or type of anastomosis (hand - sewn vs stapled) has similar postoperative outcomes and bowel function.
- One-third of patients undergoing restorative proctocolectomy (RPC) develop at least one adverse event, which include anastomotic separation, pelvic sepsis, anastomotic stricture, fistula, small bowel obstruction, bleeding or pouchitis.
- Ileal pouch failure is a time-dependent outcome with an estimated 5–10% pouch excision rate or indefinite diversion at 10 years following RPC.
- In patients who are threatened with pouch failure salvage surgery may be considered, with successful outcome in 75% of patients.

Radiation proctitis

The pathology of this condition is covered by the section on radiation enteropathy in the small bowel chapter. Clearly, radiation proctitis only occurs when the intact rectum is irradiated in the course of pelvic radiotherapy. This is usually delivered for the treatment of cervical, bladder or prostatic cancer. However, with increasing interest in the local treatment of rectal cancer, radiation proctitis can be a problem, especially when local excision is combined with external beam therapy. In addition, if pre- or postoperative radiotherapy is employed with anterior resection, the rectal remnant or (in the case of postoperative treatment) the colon used in the reconstruction can be affected. The main symptoms are rectal bleeding and diarrhoea, but in severe cases stricturing or fistulation into the vagina or bladder may occur. In extreme situations, a diversion colostomy or even rectal excision may be required.

Diversion colitis

When a section of the large bowel is isolated from the rest of the colon (usually the rectum and distal sigmoid after a Hartmann's procedure) the mucosa may become inflamed, causing mucus discharge, rectal bleeding and tenesmus. Histologically, this is indistinguishable from ulcerative colitis, but is thought to be due to a lack of intraluminal short-chain fatty acids. There is some evidence that the condition responds to butyrate-containing enemas.

Neutropenic colitis

This may occur as a complication of chemotherapy (especially for leukaemia or lymphoma), but it may also occur in other neutropenic patients and it has been reported in patients with large bowel and terminal ileal carcinoma. It is caused by superinfection by *C. septicum* or other clostridial species (*Clostridium sphenoides, C. perfringens, C. sordellii,* etc.), with the development of gas gangrene of the bowel and severe toxaemia. Fulminating neutropenic colitis carries a mortality of close to 100% as the only effective treatment is colectomy, and this is usually contraindicated by the underlying condition.

Collagenous colitis

Histologically, this disease is characterized by debasement of the epithelium, loss of goblet cells, neutrophil infiltration and subepithelial deposition of collagen which forms an irregular layer 7–25 mm in thickness. Its aetiology is unknown, and it presents with diarrhoea. The symptoms may respond to sulfasalazine, but in severe cases steroid therapy is required.

Infections and infestations of the large bowel

Amoebic dysentery

Amoebic dysentery is caused by *Entamoeba histolytica* which exists either as an active motile trophozoite or as a cyst. The former is characteristic of amoebic dysentery whereas the latter is more hardy and occurs in formed stool. The cysts are responsible for transmission of the infection. Amoebiasis is a common cause of diarrhoea in warm and humid parts of the world but sporadic cases may also occur in Europe and the USA. The amoebae cause colonic ulceration with the intervening mucosa being normal or hyperaemic.

Clinically the symptoms may be similar to ulcerative colitis and vary from mild diarrhoea to severe bloody diarrhoea. Metastatic amoebic infection may occur throughout the body with liver abscesses being the most common. These occur in 1–3% of cases and present with fever, pain and tenderness in the right upper quadrant. Alternatively they may present with a pyrexia of unknown origin. Amoebic dysentery may also be complicated by colonic perforation, stricture formation or severe haemorrhage.

Definitive diagnosis is made by stool microscopy, rectal biopsy or serology. The stool is examined for trophozoites and cysts but the former are difficult to find unless fresh specimens are examined. Rectal biopsy is useful as amoebae may be seen in the surface exudate or within the lymphatics. Serology is most valuable in patients who have amoebiasis outside the intestine when the stool tests are often negative. Serology is also useful for patients with suspected IBD in order to exclude amoebiasis. The drug of choice is metronidazole for acute amoebic dysentery whereas diloxanide furoate is effective in chronic intestinal infections associated with cysts in the stool.

Bacillary dysentery

This is due to infection with *Shigella* organisms and four main serologically distinct groups are implicated in human disease: *Shigella dysenteriae, S. flexneri, S. boydii* and *S. sonnei*. The disease is spread by person-to-person contact, as faecal excretion of the bacterium continues for 1–4 weeks in untreated individuals. The organism invades the colonic wall leading to ulceration and the clinical features are characterized by fever, abdominal pain and watery diarrhoea leading to passage of blood and mucus. Complications include haemolytic uraemic syndrome, aseptic meningitis and Reiter syndrome (arthritis, uveitis and conjunctivitis). This last complication is associated with HLA-B27. The diagnosis is made by stool culture and serotyping for identification of the specific species. The disease is self-limiting and is usually quite mild but in severe cases ampicillin or co-trimoxazole should be used. It may also be necessary to apply specific rehydration therapy, preferably orally although intravenous fluids are sometimes required.

Salmonellosis

Salmonella may cause simple gastroenteritis, which is common in the developed world, or typhoid, which is common in areas with poor sanitation. *Salmonella* gastroenteritis is caused by *Salmonella typhimurium* and *S. enteritidis* and the major non-human reservoir of these organisms is in poultry and domestic livestock. Typhoid on the other hand is caused by *S. typhi* and *S. paratyphi* and humans are the only important reservoir for these organisms. *Salmonella* infections involve both the small intestine and the large intestine. The three major patterns of *Salmonella* infection are:

- acute gastroenteritis
- typhoid (enteric) fever
- asymptomatic carrier state.

Salmonella gastroenteritis

This is relatively common with about 500 cases per 1 000 000 of the population occurring each year in the UK. The symptoms of gastroenteritis including headache, abdominal pain, fever and diarrhoea develop between 8 and 48 hours after eating contaminated food. The stool is watery and very rarely contains blood. The diagnosis can be made by culture of the stool. The disease is self-limiting and antibiotics are not valuable unless the patient is very ill with bacteraemia. In severe cases patients may require to be hospitalized and resuscitated with intravenous fluids.

Typhoid fever

Here the incubation period is in the region of 10–20 days and the clinical onset consists of progressive fever, vague abdominal pain, headache and cough. After about 5 days the patient develops abdominal distension, diarrhoea and splenomegaly. A cutaneous vasculitis occurs mainly on the abdomen giving rise to characteristic 'rose' spots. The disease may be complicated by intestinal haemorrhage and perforation and systemic problems including meningitis, encephalomyelitis, disseminated intravascular coagulation, hepatitis, pancreatitis and ectopic *Salmonella* infections (e.g. in bone).

Diagnosis is made by culture of the organism from blood, stool, urine or bone marrow. Blood cultures are positive in 80% of patients in the first week of symptoms, whereas the stool cultures are usually positive during the second and third weeks. Serological tests can be useful when culture facilities are not available and a rising titre of O antibody indicates active infection. Typhoid fever should be treated with antibiotics and the most effective drug is chloramphenicol. However, drug-resistant organisms are becoming common and it may be necessary to use amoxicillin or co-trimoxazole. When perforation occurs laparotomy with peritoneal cleansing and simple closure of the perforation is necessary. This must be supplemented with antibiotic therapy (see also Chapter 29).

Asymptomatic carriers

Carriers of *S. typhi* are important to identify as they represent a major source of the disease. The bacterium is commonly carried in the gallbladder and cholecystectomy may be necessary. Recently, however, ciprofloxacin, with its good biliary penetration, has been shown to be reasonably effective in eliminating the carrier state.

Escherichia coli infections

E. coli is a normal commensal of the gastrointestinal tract of human beings and most strains are harmless. Some strains however cause gastroenteritis. These are divided into four main groups:

- enteropathogenic
- enterotoxigenic
- enteroinvasive
- enterohaemorrhagic.

Pathogenic strains of *E. coli* can be spread by the ingestion of contaminated food (often cold meat) and by person-to-person transmission. Often both types of spread can be implicated in an epidemic with contaminated food being the initiating factor and a second wave of infection being caused by person-to-person transmission. The most common pathogenic serotype is 0157:H7, which produces cytotoxins and may lead to the haemolytic uraemic syndrome. This carries a high mortality, particularly in elderly individuals. Diagnosis of *E. coli* infection is made by the isolation of the pathogenic strain from the stool but tests for verotoxin can be more sensitive than culture.

Campylobacter infection

In man, *Campylobacter* infections are usually the result of ingestion of infected food, and poultry in particular. However, they may also be water borne, milk related and contracted from dogs. *Campylobacter* species affect both the small and large intestine and most strains produce an enterotoxin and one or more cytotoxins. Clinically there is usually a self-limiting episode of diarrhoea which lasts for a few days, and this may be preceded by fever, headache, myalgic malaise and abdominal pain. Rectal bleeding often follows the diarrhoea. Diagnosis is made by stool culture and the bacterium may be present in the stool for up to 5 weeks. Occasionally, infection may be complicated by massive lower gastrointestinal haemorrhage, Reiter syndrome and Guillain–Barré syndrome. As the majority of infections are self-limiting, treatment is not usually necessary but, in a prolonged illness, antibiotics such as erythromycin, tetracycline, chloramphenicol and gentamicin are appropriate.

Clostridium difficile infection

C. difficile is a large Gram-positive anaerobic bacillus which produces two toxins, A and B. A is directly cytotoxic to colonic epithelial cells but the role of B is not completely clear. *C. difficile* infection (CDI) is important in both antibiotic-associated diarrhoea and the more serious pseudomembranous colitis. There has been a significant increase in CDI in most Western countries and large outbreaks of CDI-associated colitis have been caused by the PCR ribotype O27 CD strain (CD027). Epidemiological studies have indicated that CD027 is most frequently acquired in the community and emerges sporadically in hospital patients following administration of antibiotics.

Pseudomembranous colitis is a form of colitis in which a pseudomembrane is caused by surface inflammatory exudate. Both mild CDI and pseudomembranous colitis are associated with the use of almost all antibiotics, although lincomycin, clindamycin, ampicillin and amoxicillin account for about 80% of all cases. The pathogenesis of both forms of CDI is unclear. The organism is present under normal circumstances in the stools of about 2% of the population but it may be present in small numbers in a much larger proportion. The factors which favour proliferation of *C. difficile* in the colon are unclear but the administration of antibiotics is clearly of major importance. General practitioners and hospital doctors are now under strict guidelines to avoid prescription of unnecessary antibiotics and hospitals have antibiotic policies which indicate the antibiotics which can be used in patients that need this treatment.

Strict personal hygiene, such as washing hands after going to the toilet, can reduce the spread of these infections. Good cleaning practices and strict hygiene measures in hospitals help to prevent contamination of equipment and personnel with bacteria and spores as *C. difficile* is very contagious and it can spread very easily through fomites. Other suggested contributory factors include proton pump inhibitor administration.

Clinically, the disease varies from a mild illness with a moderate increase in bowel frequency to severe bloody diarrhoea associated with marked abdominal pain and tenesmus. Fever is present in the majority of cases, and on examination there may be abdominal tenderness and peritonism. While the disease usually begins during a course of antibiotics, in about 30% of cases symptoms may start after the antibiotic has been discontinued. CDI may occur and complicate patients suffering from ulcerative colitis.

Diagnosis

The diagnosis is made by both culture of *C. difficile* from stool and from the demonstration of the toxin (usually B) in the stool. Most laboratories perform toxin tests based on an enzyme immunoassay (EIA) method. These tests have sensitivities ranging from 60% to 70% and specificities of 98%. Symptomatic patients with negative tests should be tested by other more sensitive tests: cell culture cytotoxicity neutralization assays (CCNAs). Toxigenic culture has emerged as the new gold standard but because these assays detect a gene that encodes toxin and not the toxin itself, it is important that these tests are performed only in patients with diarrhoea. These more recent molecular assays have been shown to be superior to toxin EIAs and CCNA but not to toxigenic culture.

The diagnosis of pseudomembranous colitis depends on endoscopic appearances which consist of exudative punctate raised plaques with intervening areas of oedematous mucosa. Biopsy shows areas of focal necrosis with polymorphonuclear leucocytes forming characteristic 'summit' lesions.

Treatment

Mild disease

The first measure in the treatment of patients with CDI is to discontinue the antibiotic responsible for the disease to enable the normal colonic flora to recover. This measure alone has been shown to be effective in 15–25% of patients. Fluid and electrolyte replacement may be necessary in these patients. Patients with mild CDI (diarrhoea with minimal symptoms) are monitored for 48 hours for symptomatic improvement before treatment with an antibiotic is instituted. Patients with more serious infection (high fever, pronounced leucocytosis, severe abdominal pain and absence of diarrhoea) require early institution of the appropriate antibiotic therapy.

Antibiotic therapy

Vancomycin and metronidazole are the two primary antibiotics used in the treatment of CDI in 7–10 day courses. Oral metronidazole is the preferred therapy as it is readily absorbed in the upper gastrointestinal tract, and usually well tolerated (250–500 mg administered four times daily or 500–750 mg three times daily). In patients who are intolerant to oral metronidazole, the antibiotic is administered intravenously (500–750 mg three or four times a day). Side effects of metronidazole include an unpleasant metallic taste, nausea, vomiting, diarrhoea, abdominal pain, headache, pruritus, erythematous rashes, dizziness and reversible neutropenia. Vancomycin is usually reserved for severe, life-threatening CDI, for patients unable to tolerate metronidazole or for patients without symptom resolution after completing a course of metronidazole. The recent emergence of vancomycin-resistant organisms is a growing concern. Oral vancomycin, which is not appreciably absorbed and is excreted in the stool unchanged, administered in oral doses of either 125 mg or 500 mg four times daily in adults are recommended. The use of a rectal vancomycin enema (500 mg diluted in 1000 mL of 0.9% sodium chloride injection) is an alternative. Several randomized clinical studies have not shown significant differences between the results of metronidazole and vancomycin therapy in the treatment of CDI. In a recent Cochrane review, Nelson concluded that one should choose the antibiotic that brings both symptomatic cure and bacteriological cure. In this regard, teicoplanin appears to be the best choice because the available evidence suggests that it is better than vancomycin for bacteriological cure and has borderline superior effectiveness in terms of symptomatic cure. However, this antibiotic therapy is currently not available in the USA.

Approximately 20–50% of patients have a relapse of diarrhoea from CDI after an initial course of antibiotic therapy. Recurrent episodes respond to the same 10 day course of antibiotic therapy used in the first episode.

Alternative therapies

Alternative antibiotics for CDI include bacitracin and teicoplanin. Binding resins such as cholestyramine or colestipol are less effective.

Faecal microbiota transplantation (FMT)

FMT (previously known as faecal bacteriotherapy) is increasingly used in the treatment of recurrent disease. Faecal transplantation is performed at colonoscopy by infusing fresh donor faeces into caecum. The patients have a whole-bowel lavage with polyethylene glycol solution prior to FMT. FMT has also been reported to benefit other colonic disorders, e.g. constipation and IBS. Intravenous immunoglobulin is used in some centres for severe disease but evidence for the efficacy of this treatment remains uncertain.

Severe disease

There is no consensus definition for severe CDI, nor is there agreement on the most important clinical indicators that differentiate disease severity. Thus various systems are used.

- Traditional indices of severe CDI: white blood cell (WBC) count >15 000 cells/μL or a serum creatinine level ≥1.5 times the premorbid level.
- Indices based on the Quebec outbreak: WBC count >20 000 cells/μL and elevated serum creatinine – significantly elevated WBC counts in the absence of any other cause should raise suspicion for CDI.
- Scoring system used in clinical trials to identify patients with severe infection – one point each is given for age >60 years, temperature >38.3°C, serum albumin <2.5 mg/dL (25 g/L), or peripheral white blood cell count >15 000 cells/μL within 48 hours of enrolment to study. Two points are given for endoscopic evidence of pseudomembranous colitis or treatment in the intensive care unit. Patients with two or more points are considered to have severe disease.
- Severe disease defined as ≥10 bowel movements per day, a peripheral WBC count ≥20 000 cells/μL or severe abdominal pain.

The reported incidence of severe CDI varies widely given the lack of a consensus definition. The risk of complications during first CDI recurrence in the Quebec outbreak caused by the hypervirulent North American Pulsed Field type 1 (NAP1) strain was 11%. Complications included shock, need for colectomy, megacolon, perforation or death within 30 days. Older age, high leucocyte count and acute renal failure were strongly associated with a complicated course.

Treatment

Patients with severe CDI should receive antibiotic therapy, supportive care and close monitoring in ICU. Surgery should be considered if the patient's clinical status fails to improve and the serum lactate rises above 2.2. Toxic megacolon should be suspected if the patient develops abdominal distension with diminution of diarrhoea (reflects onset of paralytic ileus resulting from loss of colonic muscular tone)

Antibiotics

Oral vancomycin is the first-line therapy for severe CDI. Usually oral vancomycin is administered in doses of 125 mg four times daily in severely ill patients, as reports indicate that levels achieved with higher doses (500 mg four times daily) are equivalent. Patients with severe disease who do not demonstrate clinical improvement with oral vancomycin are changed to fidaxomicin. If this is not available, oral metronidazole (500 mg three times daily or 250 mg four times daily for 14 days) is used. In severely ill patients with paralytic ileus, intravenous metronidazole in a dose of 500 mg every 8 hours is used. Intravenous tigecycline has been used in a small number of patients with severe CDI that was refractory to standard therapy.

Intracolonic antibiotics

Vancomycin enema is an effective adjunctive therapy in patients who cannot tolerate the oral preparation, or patients who have megacolon or ileus. In a retrospective case series of nine patients with refractory symptoms, toxic megacolon or fulminant colitis, rectal vancomycin was administered in addition to standard antibiotics, eight patients had complete resolution of symptoms and one patient died from multisystem organ failure. Thus intracolonic vancomycin is particularly useful in patients with profound ileus. Although the optimal dosing and volume has not been established, rectal vancomycin is often given as a retention enema containing 500 mg in 100 mL of normal saline every 6 hours. In addition some reports suggest that patients with megacolon may benefit from colonoscopic decompression and placement of a tube in the right colon which can be perfused with a 1 mg/mL solution of vancomycin in normal saline to deliver a total dose of 1–2 g per day. According to one report, this is safely achieved by the creation of a laparoscopic terminal loop ileostomy with insertion of a Malecot catheter in the distal limb. Dose adjustments may be required depending on individual circumstances including extent of colonic disease and patient weight. The promising results with this laparoscopic approach require confirmation by larger studies. It is important to note that vancomycin can be absorbed through inflamed colonic mucosa and cause systemic toxicity.

Surgery

Emergency surgery is required for toxic megacolon, perforation or impending perforation, necrotizing colitis or rapidly progressive and/or refractory disease with systemic inflammatory response syndrome leading to multiorgan system failure. Although the optimal timing of surgery remains uncertain, the emphasis in recent years is towards earlier surgical intervention; the appropriate operation is subtotal colectomy with preservation of the rectum. In a retrospective review, colectomy was most beneficial for immunocompetent patients aged ≥65 years with a WBC count ≥20 000 cells/μL and/or a plasma lactate between 2.2 and 4.9 meq/L. The reported mortality of emergency surgery for complicated severe CDI is 40–50%.

Chlamydial infection

The commonest type of chlamydial infection is a sexually transmitted urethritis but the rectum may also be affected, particularly in men who have sex with men. This leads to an ulcerative proctitis characterized by rectal bleeding. The diagnosis depends on isolation of the organism from the rectum and treatment with tetracycline is usually effective.

Schistosomiasis

It should be remembered that the adult worms of *Schistosoma mansoni* reside mainly in the inferior mesenteric vein and *S. japonicum* in the supermesenteric vein. These worms produce large numbers of eggs, which are retained in the wall of the intestine where they cause granulomatous reactions and some penetrate the intestinal wall and are passed with the faeces. The presence of the eggs in the intestinal wall may give rise to rectal bleeding and occasionally large polypoid lesions may occur and can be misdiagnosed as cancers, usually in the descending and sigmoid colon.

Cryptosporidiosis

The parasite *Cryptosporidium* causes diarrhoea and may be fatal in immunocompromised patients, particularly those with AIDS. The parasite infects the enterocytes of both the small and large intestine and transmission appears to be either person to person or caused by ingestion of infected water. There is no known effective treatment although in normal individuals the gastroenteritis tends to be mild and self-limiting. In patients with AIDS, however, *Cryptosporidium* causes protracted watery diarrhoea with severe weight loss.

Colorectal cancer

Epidemiology

There are about 140 000 and 40 000 new cases of colorectal cancer diagnosed each year in the USA and UK respectively. Around two-thirds are located in the colon and one-third in the rectum. This equates to 55 new cases per 100 000 population in the USA and 55.9 in the UK. It is the second most common cancer in the UK (Cancer UK) and the USA (http://www.cancer.org/downloads/STT/F861708_finalforweb.pdf). The incidence of colorectal cancer is higher in males and black people. It has been also shown to increase with age, with the majority of the patients being diagnosed at the age of 50 or above. In recent years, however, there has been an increase in the number of younger patients diagnosed with the disease, raising more questions about the already established screening programmes.

Cancer of the colon and rectum is one of the few cancers that have been associated strongly with a number of environmental factors. Consequently, this has been responsible for the significant variability among different populations, and thus the higher incidence in the Western world. Furthermore, there are significant variations among subgroups of the population. In the UK, the incidence in England is significantly lower than in Scotland. The incidence in males ranges from 57 per 100 000 in England compared with 68 per 100 000 in Scotland. The pattern prevails in females (37 per 100 000 females diagnosed with colorectal cancer in England compared with 44 per 100 000 in Scotland) (Cancer UK report).

Aetiology and associated risk factors

The colorectal cancer risk is influenced by both environmental and genetic factors, with the lifetime risk up to 5%. There are a number of studies investigating the aetiology of colorectal cancer, but the exact cause is still not fully clear. There are a number of environmental and genetic factors that have been strongly associated with the disease and therefore are considered to play a role in the carcinogenesis. The exact mechanism is not fully understood yet.

Environmental factors

Diet is one of the major environmental risk factors that have been associated with colorectal cancer. It has been suggested that the Western diet is a strong risk factor (WHO Cancer report). The consumption of red meat and poultry was shown to be associated with colorectal cancer and might be a result of the exposure to carcinogens that can be created when cooking. The role of red meat was also assessed by three meta-analyses that confirmed the above statement. One study demonstrated a link to colorectal cancer when the meat was well cooked. This study suggested that carcinogens formed on the surface of well-cooked red meat might be responsible for this link, particularly for rectal cancer. Myoglobin has also been suggested to play a part in the carcinogenesis.

There are a vast number of publications debating the value and role of dietary fibre in colorectal cancer. There was initially a negative association of dietary fibre intake to colorectal cancer. The majority of the published studies showed that lower intake of vegetables and therefore fibre was associated with a lower risk of colon cancer. However, this claim was soon challenged by two meta-analyses that confirmed the benefits of a high intake of vegetables in terms of reducing the risk of colorectal cancer. Trock *et al.* demonstrated a reduction of risk by 40%, and Howe *et al.* by 50%. More recent prospective studies have been performed but did not demonstrate significant association of high-fibre intake as protective factors for cancer.

Micronutrients and vitamins have also been investigated. A study investigating the level of folate in serum showed that low serum levels were associated with higher rates of colorectal cancer. One meta-analysis showed that the risk was higher for colorectal cancer when the serum levels of vitamin D were reduced. Calcium and aspirin have been shown to reduce the risk for colorectal adenoma formation, thus reducing the risk of colorectal cancer.

Obesity is associated with a higher risk of colorectal cancer. Obesity is shown to increase the risk by around 25% in overweight males and 50% in obese males. The same study showed a non-statistically significant risk of 9% for obese females. Larger waist size was also linked to higher risk for both males and females. Physical activity has been shown to play an important role. High levels of physical activity have been shown to substantially reduce the risk, especially in males by almost 25%. Physical inactivity has been suggested to be responsible for up to 14% of the new colon cancer cases in the Western countries. The findings of the European Prospective Investigation in Cancer (EPIC) study confirmed these findings. This study also evaluated the added effect of these two factors and showed that the risk ratio of colon cancer was 0.38 in males who were both lean and active compared with patients who were inactive with a BMI >30.

Smoking and alcohol have also been investigated thoroughly and shown to increase the risk of colorectal cancer for both men and women. A meta-analysis published in 2009 demonstrated a statistically significant increase in risk of colorectal cancer (risk ratio = 1.20; $p < 0.0001$), especially in male smokers (risk ration = 1.38, $p < 0.0001$). The same study showed a dose-dependent risk. The relative risk for male heavy smokers increased to 1.28 from 1.19, which was the risk for patients who had consumed fewer than 30 pack years. Regular alcohol consumption has a dose-dependent risk of colorectal cancer. This can increase by up to 50% for an intake of 50 g/day of alcohol or above compared with occasional drinkers.

Familial and genetic factors

Family history of colorectal cancer is known to increase the risk of colorectal cancer. The risk doubles when there is a history of a first-degree family member diagnosed with colorectal cancer. The risk is even higher when the relatives were diagnosed with the disease at a younger age (younger than 50) and for patients with more than one relative with the disease.

FAP and the hereditary non-polyposis colorectal cancer (HNPCC) have been linked with about 5% of the diagnosed colorectal cancers. The presence of early multiple bowel adenomas is the distinctive feature of FAP, which is responsible for about 1% of colorectal cancers. The presence of FAP increases the risk of developing colorectal cancer to 100% by the age of 40. It is an autosomal dominant disorder, caused by a germline mutation in the adenomatous polyposis coli (*APC*) gene, which is a tumour suppressor gene and found on chromosome 5q21. Early onset of colorectal cancer is typical for HNPCC, which is estimated to be responsible for 1–4% of colon cancers. HNPCC is secondary to germline DNA mismatch repair gene mutations in the majority of the cases. The overall lifetime cancer risk of developing colorectal cancer in patients with these mutations by the age of 70 is 91% for men and 61% for women. The same study demonstrated that the risk was significantly higher for males (74%) than for females (30%; $p = 0.006$).

Other, rarer syndromes have also been shown to increase the risk of developing colorectal cancer. Peutz – Jeghers syndrome is an autosomal dominant condition characterized by the presence of hamartomatous polyps in the small bowel and mucocutaneous melanin pigmentation of the perioral region, hands and feet. Patients with the syndrome have 39% higher risk of developing colorectal cancer by the age of 70. MYH (mutY homolog gene mutation)-associated polyposis and juvenile polyposis, both autosomal dominant conditions, have been associated with 35–53% and 17–68% increased risk of colorectal cancer respectively.

Inflammatory bowel disease

Patients diagnosed with IBD carry higher risk of colorectal cancer than the general population. This risk has been extensively investigated and is well established. The risk of colorectal cancer in ulcerative colitis patients is more pronounced. Colorectal cancer on the background of chronic ulcerative colitis has been described as being more aggressive and with poorer prognosis than the general population. The prevalence of colorectal cancer in patients with ulcerative colitis has been reported up to 3.7%. The cumulative risk of developing colorectal cancer has been reported to increase by 1% per year after 10 years from the date of onset of active ulcerative colitis. The relative risk of developing colorectal cancer on the background of ulcerative colitis at an early age has been reported up to 38-fold, with reports of 14.8 and 2.8 for patients with pancolitis and left-sided colitis respectively. The risk of developing colorectal cancer in ulcerative colitis has also been found to be significantly increased in patients with total colitis.

Earlier studies showed that the incidence of colorectal cancer in patients with CD was higher cancer, with other studies showing that CD-associated colorectal cancer had a poorer prognosis than sporadic colorectal cancer. The risk ratio for developing colorectal cancer on the background of CD is 4.5 in patients with colonic disease, compared with the general population, with a cumulative risk of 2.9% at 10 years.

Clinical presentation

Initially patients with colorectal cancer may be asymptomatic. As the disease progresses, development of systemic symptoms and symptoms specific to the location of the cancer may be observed. Rectal bleeding is one of the commonest presentations of colorectal cancer. However, it may be the result of a benign pathology and therefore has low positive predictive value for colorectal cancer. The prevalence, in populations of over 20 years old, ranges between 15.5% and 24%. A recent meta-analysis of 13 studies with 18 634 patients showed a positive predictive value (PPV) of 5.3% that increased to 8.1% in a subgroup analysis of studies that included patients aged 50 or more. Heterogeneity was moderate. A study by Fine *et al.* suggested that the colour of bleeding reported by patients was inaccurate because of misinterpretation and the use of a colour card was proposed in order to help patients to be more accurate. The same study reported that the majority of bright red bleeding (about 83%) was linked to distal (60 cm) pathology. However, more proximal pathology was identified in 20 of 217 (9.22%) patients. A community-based trial showed that colorectal cancer was more likely when the blood was mixed with faeces and when it was reported as being dark (likelihood ratio 3.0). The picture is slightly different in secondary care, with the positive predictive value for colorectal cancer being higher for patients with a history of dark and bright blood (13.2%) than dark blood (10.6%) or bright blood alone (4.3%), and for when the blood was mixed with faeces (11.0%) than when separated from stools (3.4%). One study of 226 patients evaluated the risk of rectal bleeding in the presence of perianal symptoms and showed a negative likelihood ratio of 2.90, with poor sensitivity and specificity. This suggests that rectal bleeding in the presence of perianal symptoms carries low risk of colorectal cancer.

Patients with colorectal cancer often present with a ‘change in bowel habit’. The definition in the UK refers to the passage of loose stools and/or increased frequency of defecation that persists for more than 6 weeks. It has been reported by a number of studies that change in bowel habit may occur in up to 91% of patients with distal and 61% of patients with proximal cancer. Constipation is not currently considered as a risk factor; recent studies suggest that there may be a link and therefore should be used to refer patients for further investigation. A meta-analysis showed a positive likelihood ratio of 1.8 and negative likelihood ratio of 0.7 when it is associated with rectal bleeding.

Unintentional weight loss has been regarded as a predicting factor of colorectal cancer. However there has been no conclusion about the percentage of weight loss that should be considered clinically relevant. Loss of 5% of weight in a period between 6 and 12 months is generally considered significant. In colorectal cancer, weight loss is usually a late presentation with a median of 27 (9–42) weeks. The same study demonstrated a higher incidence of weight loss for proximal (46%) than distal (34%) tumours, but did not achieve statistical significance. The positive predictive value for colorectal cancer in patients with weight loss increases to 4.7% when it is associated with rectal bleeding, to 3.1% with diarrhoea and 3% with constipation. A recent meta-analysis of six studies that included 1468 patients demonstrated that the combination of weight loss and rectal bleeding increased the positive likelihood ratio to 1.88 and negative likelihood ratio to 0.93. There was low heterogeneity among the included studies.

Abdominal pain may be the first presentation. Up to 30% of the general population experience abdominal pain every year. This is usually a result of a benign condition. It has low sensitivity and specificity when it is the sole symptom. A recent meta-analysis demonstrated a positive likelihood ratio of 2.47 and negative likelihood ratio of 0.75. The same study showed low sensitivity (31%) but high specificity (91%). The heterogeneity between the studies was high though. The positive predictive value increased to 7.6% when abdominal pain was associated with rectal bleeding but the between-study heterogeneity remained high. Hamilton *et al.* demonstrated that there was an increased PPV of 3.4% when it was associated with weight loss, 3.1% with rectal bleeding and 1.9% with diarrhoea.

Iron-deficiency anaemia is present in 11–57% of colorectal cancers and is more common for right-sided tumours. The patients may present with fatigue and/or shortness of breath. In the UK, haemoglobin of <11 for females and 12 g/dL for males indicates the need for further investigations. A meta-analysis of four studies (928 patients) assessed the PPV of anaemia and found it to be high, with high heterogeneity mainly due to the significant number of true positives reported from one study. A subgroup analysis was performed excluding that study and showed a PPV of 7% with the heterogeneity between the studies reduced.

Investigation

Clinical examination

A full history and examination is essential before any management plan is made. History should also include questions assessing the patient's performance status and comorbidities, as this information may be pivotal to the management plan. During clinical examination the patient should be specifically inspected for signs of anaemia, jaundice (suggestive of liver metastases) and malaise, and examined for abdominal distension, ascites, masses and abdominal tenderness. Digital rectal examination is essential, as it can provide essential information at the initial stage if the tumour is within the lower rectum.

Blood tests

Routine blood tests are performed including a full blood count, urea and electrolytes, liver function tests and coagulation. More specifically, the haemoglobin is measured to confirm or rule out anaemia, the liver function tests and coagulation to assess the liver function and raise any question of liver metastases.

Tumour markers, such as the carcinoembryonic antigen (CEA) and cancer antigen (CA) 19-9 are measured. CEA is an oncofetal antigen discovered in 1968. It is a product of the normal fetal gut tissue and epithelial tumours, especially those of the large bowel. It can, however, increase in smokers, IBD, pancreatitis, liver disease and in patients with epithelial tumours at other sites. CA 19-9 was discovered in patients with colon and pancreatic cancer. Negative CEA and CA 19-19, with a cut-off of 5 ng/mL and 37 U/mL, respectively, have been shown to significantly improve the overall survival following potentially curative operation. Care should be taken with mucin-producing and poorly differentiated adenocarcinomas of the colon and rectum which do not often present with a raised CEA and are classed as high-risk tumours for local failure and poor long-term prognosis.

Proctoscopy and rigid sigmoidoscopy

Proctoscopy and/or rigid sigmoidoscopy can be performed in an outpatient setting, confirming a lower rectal tumour and providing at the same time the opportunity to get biopsies from the tumour and therefore plan the patient's further management more efficiently. In addition, it may reveal a different pathology (i.e. IBD) or the presence of streaks of blood in the lumen of the rectum, which is strongly indicative of pathology in the sigmoid above the reach of the rigid rectoscope.

Flexible sigmoidoscopy

Flexible sigmoidoscopy can be performed relatively quickly and is virtually risk free. Therefore, the National Institute for Health and Clinical Excellence (NICE) in the UK suggests it to be the most appropriate initial investigation for the majority of symptomatic patients, especially the patients with symptoms suggesting possible left-sided lesions. Flexible sigmoidoscope can only reach up to 60 cm. When there are no positive findings in patients without any right-sided symptoms, the likelihood of right-sided cancer is low.

Colonoscopy

NICE advises the use of colonoscopy in patients who are considered at high risk of developing colon cancer because of older age, a clinically palpable abdominal mass, iron-deficiency anaemia or symptoms such as abdominal pain with loss of appetite and weight. Colonoscopy is also necessary when there is significant clinical doubt following a flexible sigmoidoscopy. Diagnostic colonoscopy is usually appropriate for patients with right-sided symptoms, except for those with palpable masses, for whom imaging may be more appropriate (barium enema or CT). Colonoscopy should also be considered when a rectal tumour is diagnosed in order to exclude any synchronous tumours. About 4% of patients diagnosed with primary colorectal cancer will have synchronous colon cancers. Colonoscopy enables direct visualization of the tumour and its extent with assessment of its fixity to the surrounding tissues; and its potential for complete obstruction. In addition, biopsies can be taken and any significant active bleeding can be controlled.

Imaging

Accurate identification of metastatic disease is key for the decision to operate on a patient. CT colonography and barium enema can be used instead of colonoscopy for the initial diagnosis of the tumour. For colon cancer, a CT scan is used to stage local disease and assess for potential distant metastases (i.e. lung, liver). CT of the chest, abdomen and pelvis is usually requested. For rectal cancers MRI is used to stage local disease and CT to assess potential distant metastases. There is not enough evidence yet to support the regular use of positron emission tomography (PET) or PET/CT scans in this group of patients. It is therefore employed when clinical and radiological findings from other scans are equivocal. For low rectal cancers endorectal ultrasound has been used with high accuracy for the early stage cancers.

CT colonography

CT colonography is a modern application of the conventional CT scan that can provide information on the entire colon. It is minimally invasive and has been reported to be easier to tolerate by the patients. It is required that the patients undergo bowel preparation before the procedure, although more recent studies have been investigating the potential to avoid this. The side effect of this modality is the increased radiation that patients receive compared with conventional colonoscopy. Low radiation protocols are currently being investigated, with early results showing that it is feasible to provide good images. There have been three large multicentre studies published with significant differences reported in terms of the accuracy of CT colonography. Pickhardt *et al.* found comparable sensitivities between CT colonography (86%) and colonoscopy (90%) in the detection of polyps of ≥6 mm.

However, subsequent publications showed that colonoscopy was significantly superior to CT colonography by detecting up to 99% of lesions/polyps. A more recent study showed that CT colonography has high sensitivity (97%) and specificity (about 91%) and that it can be more accurate than barium enema studies, but less accurate than colonoscopy. A meta-analysis of 49 studies of 11 151 patients was performed to assess the sensitivity of CT colonography and colonoscopy and demonstrated high sensitivity for both tests with moderate between-study heterogeneity when both cathartic and tagging factors were combined for bowel preparation.

Barium enema

Barium enema is well established in the NHS in the UK. It is safe and readily available. Bowel preparation is necessary but there is no need to sedate the patient, and therefore the patient can be discharged home the same day. Barium enema on its own has lower sensitivity than colonoscopy, and any negative results should be interpreted carefully in association with the patient's symptoms. Another disadvantage of barium enema is the inability to take tissue biopsies or remove any polyps. For these reasons its use has been declining in NHS centres where colonoscopy or CT colonography is available.

Locoregional staging

Computed tomography

CT is extensively used in the preoperative assessment of patients with colorectal cancer. It is used to assess local and distant disease. Studies have shown variability in the assessment of the extent of local disease. Modern CT scanners can also provide detailed reconstructed three-dimensional images that can help to improve the interpretation of two-dimensional images. A recent study has demonstrated that modern CT scanners are highly accurate in predicting the pathological T stage (pT). The diagnostic accuracy for T2 and T3 was 94.1%, and 100% for T4 tumours. The same study demonstrated high diagnostic accuracy (80.5%) for detecting lymph node metastases. In older studies, CT was less accurate in detecting and staging tumour and lymph node metastases.

A large study demonstrated that lymph node metastases were often missed by CT. This may be because of the weakness of CT to detect micrometastases and microalterations of the lymph node shape and size. In order to address this, Kanamato *et al.* analysed every lymph node that was detectable by CT and measured its longest and shortest diameter. They demonstrated that a short–long-axis diameter ratio of 0.8 or greater as an index for the diagnosis of metastatic lymph nodes achieved an accuracy index of 80% per node. As the technology evolves, it is likely that CT scanners will become even more accurate. There are radiological criteria for CT that have been established in order to facilitate radiological staging. The criteria are described in Table 30.13.

Table 30.13 CT criteria for colonic T staging

Stage	Description
T1	Intraluminal lesion without evidence of bowel wall invasion (bowel wall thickening)
T2	Thickened colon wall without disruption of muscularis propria
T3	Thickened colon wall with discrete mass extending into pericolic fat
T4a	Irregular advancing edge of tumour penetrating adjacent organs
T4b	Breach of the peritoneum that covers the colon
T4c	Perforated tumours with evidence of pericolonic gas and free fluid in the abdomen

Magnetic resonance imaging

MRI is the investigation of choice for pretreatment local (T) staging for rectal tumours. High-resolution MRI can clearly depict the bowel layers and accurately detect the depth of tumour invasion, extramural invasion and the relationship of the tumour to the surrounding anatomical structures. MRI has high diagnostic accuracy in predicting the T stage and the distance of the tumour from the circumferential margins. It correctly anticipates the subsequent histopathological examination in 85% of T3 and T4 tumours. MRI has a sensitivity of 94% and a specificity of 85% in determining the relationship between the advancing border of the tumour and the fascia propria of the rectum. Involvement of the fascia propria predicts a positive surgical circumferential margin. MRI also allows the subcategorization of T3 tumours into T3a and T3b.

Nodal staging has been always challenging as micrometastases may not be diagnosed. Nodal staging, whether assessed by ultrasound or MRI, is still unreliable, but analysis of the consistency of the node and the regularity of its capsule as demonstrated on MRI may improve its accuracy. Brown *et al.* showed that the heterogeneity of the lymph node's MRI signal was a highly specific discriminator, raising the diagnostic sensitivity to 85% and specificity of 97%. A recent study by Kim *et al.* showed the diagnostic accuracy of MRI to be 83% in detecting lymph node metastases compared with the 70% accuracy of PET. Combining both modalities increased the diagnostic accuracy to 90%. More recently, the use of nanoparticle-enhanced MRI (using ultrasmall superparamagnetic iron oxide as a contrast agent) is reported to have sensitivity in the order of 92% and specificity of around 96% for detecting lymph node metastases.

Imaging for distant metastases

Identification of distant metastases is crucial in the management of patients with colorectal cancer as it drastically changes the management of patients. CT and MRI have demonstrated high sensitivity in detecting distant metastases, providing detailed anatomical information of the affected organ and tumour extension into surrounding tissues (Figure 30.48a,b).

A meta-analysis investigating the value of ultrasound scan (USS), CT, MRI and PET in detecting liver colorectal

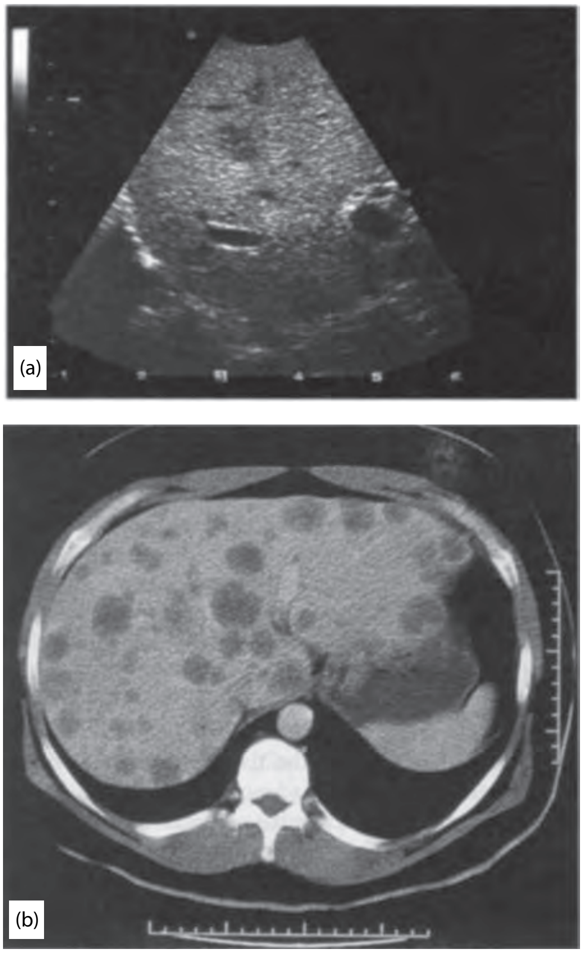


Figure 30.48 (a) Ultrasound appearance of hepatic metastases – image obtained at intraoperative contact ultrasound. (b) CT scan showing multiple hepatic metastases.

metastases demonstrated a sensitivity of 63%, 74.8%, 81.1% and 97.2% respectively and specificities of more than 93.8%, with MRI being significantly more sensitive than CT ($p=0.05$) and equally sensitive to PET ($p=0.02$). There was no significant difference in the sensitivity of PET and CT ($p>0.05$) and neither between CT and USS ($p=0.45$). Contrast-enhanced CT and gadolinium-based enhanced MRI can improve the diagnostic accuracies of CT and MRI respectively.

The incidence of lung metastases secondary to colorectal cancer (about 5% of colorectal cancers) is not as high as liver metastases. Therefore the regular use of chest CT may be questioned. In the UK, it is standard practice to perform a chest CT when staging patients with colorectal cancer. A recent study investigating the CT scans of 619 patients demonstrated high sensitivity (83.9%) and specificity (99%) for modern CT scan in detecting lung metastases. MRI is not used routinely to look for lung colorectal metastases, even though its diagnostic accuracy is similar to CT. It can be used when the findings of CT are equivocal; however, with the introduction of PET/CT, its role has been limited for staging the lungs.

PET has been demonstrated to be highly accurate in the detection of disseminated disease. A meta-analysis reported a PET sensitivity of 91% and specificity of 83% for the diagnosis of distant metastases. However, the authors admitted that only eight of 27 (29.6%) studies were of high quality, fulfilling their quality criteria at least by 80%.

Histopathological staging

In the UK Dukes' staging is still used to stage most bowel cancers. The original Dukes' classification was based entirely on pathological examination of the level of invasion in the resected specimen (Table 30.14). In spite of Dukes' concerns about the accuracy of surgeons' observations, it is now common practice to combine pathological staging with clinical assessment. Turnbull, who defined Dukes' stage D as any patient with metastatic or residual local disease, first introduced this clinicopathological staging in 1967. Others have defined Dukes' stage D as any metastatic disease in the abdomen alone or any systemic or residual local disease. In order to overcome the shortcomings of Dukes' classification, TNM systems are increasingly being applied to stage colorectal cancer patients.

TNM staging system

The TNM system was first introduced by Pierre Denoix between 1943 and 1952. In 1958 the first recommendations were published for the clinical stage classification of cancers of the breast and larynx. In 1968 the International Union Against Cancer described the classification of 23 body sites, and in 1974 and 1978 the second and third editions were published containing new site classifications. Over the years, amendments to previously published classifications have been made up to the present seventh edition of the TNM, containing rules of classification and staging that correspond with those of the American Joint Committee on Cancer's *Cancer Staging Manual*. The TNM classification system describes the anatomic extent of cancer. It is based on the fact that the choice of treatment and the chance of survival is related to the extent of the tumour at the primary site (T), the presence or absence of tumour at the regional lymph nodes (N), and the presence of metastasis beyond the regional lymph nodes.

Tumour staging can be classified prior to treatment, i.e. clinical staging (cTNM), and after resection, i.e. pathological TNM (pTNM). In cases where patients underwent preoperative chemoradiotherapy, the pathological staging is denoted by the prefix 'y' such as ypTNM. T is usually divided into four major parts (T1–T4), expressing different size or spread of the primary tumour. N and M have at least two categories each (0 or 1 indicate the absence or presence of tumour). A tumour with four degrees of T, three degrees of N and two degrees of M will have 24 TNM categories and for this purpose it is often necessary to condense these categories into a convenient number of TNM

Table 30.14 Initial Dukes' stage

Dukes' A	Invasion into but not through the bowel wall
Dukes' B	Invasion through the bowel wall but not involving lymph nodes
Dukes' C	Involvement of lymph nodes
Dukes' D	Distant metastasi(e)s

Table 30.15 TNM stage grouping for patients with colorectal cancer

Tumour	Lymph node	Metastasis
Stage 0		
Tis	N0	M0
Stage I		
T1	N0	M0
T2	N0	M0
Stage II		
T3	N0	M0
T4	N0	M0
Stage III		
Any T	N1	M0
Any T	N2	M0
Stage IV		
Any T	Any N	M1

stage groups. The purpose of the adopted grouping is to ensure that each group is more or less homogeneous with respect to survival and differs for colorectal cancer. The severity of disease, for example the extent of cancer spread (staging), has been well documented as an important determinant of long-term outcome (Table 30.15).

Medical treatment

Each individual case must be discussed in a multidisciplinary setting before any decision for medical or surgical therapy is made. Medical therapy can be in the form of neoadjuvant (preoperative), adjuvant (postoperative) or palliative, chemotherapy or chemoradiotherapy.

Neoadjuvant therapy

Neoadjuvant radiotherapy

Neoadjuvant radiotherapy or chemoradiotherapy is the most common application of preoperative medical therapy in colorectal cancer. It is mainly used for high-risk rectal cancer, when the margins are threatened or when a tumour response may change the type of surgery required to achieve complete clearance (perform an anterior resection instead of abdominoperineal resection) and preserve the anus and bowel function. Neoadjuvant chemoradiotherapy is used for T3, T4, lymph node-positive disease, and when the resection margins are threatened. For chemoradiotherapy, capecitabine and bevacizumab are the chemotherapeutic agents that are commonly used to increase the tumour's sensitivity to radiation.

Studies have shown that neoadjuvant chemoradiotherapy can reduce the local recurrence rate without changing the overall survival rates. A meta-analysis of randomized controlled studies with patients who underwent neoadjuvant therapy prior to surgery for resectable rectal cancer showed a marginal benefit for neoadjuvant radiotherapy that did not reach statistical significance. The same study showed that the cancer-specific survival and local recurrence rate were reduced in the group of neoadjuvant radiotherapy. In a more recent study the use of preoperative short-term radiotherapy reduced the 10 year local recurrence by more than 50% without an overall survival benefit.

Neoadjuvant chemotherapy

Chemotherapy prior to surgery is used to reduce the tumour's size so that it can be resected with higher likelihood of complete resection and fewer operative complications. 5-Fluorouracil (5FU) is the first-choice chemotherapy drug for colorectal cancer and is used intravenously. It can be used in combination with leucovorin (folinic acid) that has been shown to increase its action. More recently the same medication has been developed in a form that can be taken orally (capecitabine).

There are new chemotherapeutic agents that can be combined with 5FU or used alone. These include irinotecan, oxaliplating, cetuximab and penitumumab. These, apart from cetuximab, are usually used in combination with 5FU.

Preoperative chemotherapy has not been extensively used for patients in whom surgery is likely to completely remove the primary cancer. There is currently an ongoing randomized controlled study in the UK that is investigating the value of neoadjuvant chemotherapy in patients who would otherwise undergo colonic surgery and adjuvant chemotherapy (FOxTROT). This study is based on the observation that in other cancers preoperative chemotherapy has proved to be more effective than postoperative chemotherapy. This has generated the following hypotheses:

- Early optimal systemic therapy may be more effective at eradicating distant metastases than the same treatment given after the delay and immunological stress of colonic surgery.
- Shrinking the primary colon tumour before surgery may reduce the risk of incomplete surgical excision, and the risk of tumour cell shedding during surgery.

Adjuvant medical therapy

The aim of adjuvant chemotherapy is to prevent the dissemination of the disease in high-risk patients. This is discussed at the local multidisciplinary meetings with the results of the histopathology. Adjuvant chemotherapy has been shown to increase the overall survival and disease-free survival.

These are the factors that are considered to increase the risk of dissemination and used as an indication for adjuvant chemotherapy:

- positive lymph nodes
- poorly differentiated cancer cells
- vascular and perineural invasion
- T4 cancer
- evidence of localized distant metastasis during the operation.

Chemotherapy regimes that are currently in use are the following:

- FOLFOX – leucovorin, 5 Fu and oxaliplatin
- XELOX – a combination of oxaliplatin and capecitabine
- capecitabine/5 Fu
- tegafur and uracil capsules.

Palliative therapy

Radiation and chemotherapy have been used in patients who have irresectable (tumour that cannot be resected) or inoperable (tumour that is not indicated to be operated due to dissemination) disease. Radiotherapy is usually used to alleviate symptoms related to tumour expansion such as pain, lower leg oedema and lymphoedema. The aim is to shrink the tumour and reduce the effect to the local tissues. Systemic chemotherapy is used to control the dissemination of the disease and prolong the overall survival. A meta-analysis of seven trials with 866 patients showed a 35% reduction in the risk of death and 16% in the improvement of survival within a year when palliative chemotherapy was used. There was also an improvement of survival by 3.7 months (median).

Surgical treatment

Colon cancer surgery

Historical perspective

The intestinal suture was considered a 'mighty procedure in a highly vulnerable organ', and therefore a dangerous option for surgery. The formation of a stoma used to be the surgical treatment of choice for any intestinal disease and injury. Colostomy was introduced as a palliative procedure for patients with obstructive bowel cancer in 1839. Until then, patients diagnosed with obstructive cancer would have either died or been relieved by a spontaneous fistula.

Reybard of Lyons in 1833 was the first to perform a successful resection of the sigmoid colon, but he found great opposition from the Paris Academy of Medicine. The mortality following a colonic resection ranged from 60% before 1889 to 37% by 1900. Owing to the high mortality rates the staged extraperitoneal resection technique (exteriorization–resection) was usually employed.

Bowel resection with anastomosis was not widely performed until the introduction of antibiotics. The sigmoid colectomy was always performed with the formation of a defunctioning colostomy until the 1950s. At that time, the interest of the surgical community was focused on ways to reduce faecal contamination during the procedure. A number of published articles discussed the application of non-crushing clamps and the value of limited inversion of the anastomosis. Stapling techniques were reported as early as 1908, but they have found widespread application more recently.

Procedures

The aim of surgery is to achieve complete tumour resection with an adequate resection margin, along with draining the lymphatics and the creation of a tension-free anastomosis.

The following operations are performed for colon cancer:

- right colectomy
- transverse colectomy
- left colectomy
- sigmoid colectomy
- sutotal/total colectomy
- palliative resection (limited).

The majority of the new surgeons prefer to perform colon operations through a midline incision, the extent of which (upper, lower or full extent) is dictated by the type of colectomy, position of the tumour and body habitus. The advantages of a midline incision are the following:

- it facilitates quick entrance into the peritoneal cavity
- it provides adequate access to both sides of the abdomen
- it allows the formation of either a colostomy or ileostomy to either side
- the wound heals strongly and is easily closed in a single layer.

A right-sided oblique incision may be employed for right colon resections. The wound from this type of incision heals well and is possibly associated with less discomfort for the patient. However, the exposure may be compromised. This type of incision is mainly used by older surgeons who had been trained to do this technique, as previously it was more widely used. The introduction of minimally invasive surgery resulted in this technique being abandoned. Minimally invasive surgery enabled faster bowel function recovery and feeding of patients. This was associated with less discomfort and shortened hospital stay.

These results led to the introduction of minilaparotomy. Minilaparotomy has been defined as complete resection performed through a skin incision less than 7 cm long. Minilaparotomy for colon cancers was shown to be feasible, with comparable oncological outcomes, and was found to reduce hospital stay and analgesia requirements. Body habitus was found to be a crucial factor in the completion of the procedure without extending the incision.

When the peritoneal cavity is entered, it is important to check for the presence of metastatic liver and/or lymph node disease and peritoneal dissemination. The tumour is examined at the end after excluding dissemination of disease. Resectability and fixation of the tumour to the surrounding tissues are evaluated with minimum handling. Large lymph nodes can be inflammatory and, when in doubt, a frozen section can be sent for histopathological examination.

Right hemicolectomy

Right hemicolectomy is performed to resect lesions located at the caecum, ascending colon and hepatic flexure. The blood supply in this area originates from the ileocolic and right colic arteries, and therefore dictates the use of right hemicolectomy. The positioning of the surgeon is a matter of personal preference. The small bowel is retracted into the left upper quadrant of the abdominal cavity, exposing the root of the mesentery and the base of the transverse mesocolon.

The right colic and right branch of the middle colic vessels may be ligated at the beginning of the dissection. This can be more challenging in the obese patient and may not be safe to perform. The small incision in the root of the mesentery is extended to the point that the transverse colon will be divided. The mesenteric and mesocolic vessels are ligated, limiting the entire blood supply to the tumour. In more recent years, the implementation of vessel-sealing devices has changed the way that this part of the procedure is performed. These devices can securely seal small vessels (up to 7 mm diameter) within a few seconds. This technique can replace the use of clips and sutures.

The congenital peritoneal adhesions along the lateral gutter are divided using coagulation diathermy, elevating the terminal ileum and right colon from the retroperitoneal structures. This will include any lateral peritoneum involved by serosal tumour with the specimen. Care must be taken to avoid injury to the ureter, spermatic or ovarian vessels, and inferior vena cava. The ileocaecal fold of Treves is incised to prepare the terminal ileum. As the colon is elevated from the retroperitoneum, the second part of the duodenum will appear and should be displaced with care to avoid any injury. The hepatic flexure is freed from the developmental gallbladder and liver adhesions using diathermy or vessel-sealing devices. When the colon is sufficiently lifted the head of the pancreas will be visible indicating that the duodenum is adequately cleared from the field. Clamping of the blood supply can be safely performed now following development of the avascular plane around the vessels.

The lesser sac is entered by dividing the greater omentum, enabling at the same time the retraction of the posterior wall of the stomach away from the dissecting field and therefore reducing the risk of injury. The dissection of the greater omentum continues until the hepatic flexure is fully mobile. The omentum is incised at the point of the anastomosis. The bowel is therefore ready for resection. On rare occasions that the tumour invades the duodenum and the pancreas the tumour can be resected *en bloc* with the duodenum and pancreas (pancreaticoduodenectomy). This will require the employment of an expert hepatobiliary surgeon.

Surgery for transverse colon cancer

Cancer of the transverse colon can often be a challenge for the colorectal surgeon. The blood supply to this area is derived from the middle, right and left colic vessels and an anastomosis at the splenic flexure may beat risk of ischaemia, as the blood supply solely from the inferior mesenteric artery may not be sufficient. This is not usually an issue when dealing with tumours at the hepatic flexure, as there is sufficient blood supply from the ileocolic and right colic vessels.

The lymphatic drainage is also an issue when dealing with tumours of the transverse colon. Cancer in the transverse colon can metastasize to regional lymphatics through the middle colic, right colic and left colic branches. This is why some surgeons suggest subtotal colectomy as an option for this type of cancer. Other surgeons decide the type of surgery based on the location in the transverse colon, suggesting right hemicolectomy for proximal lesions and left partial or left hemicolectomy for distal lesions. For the latter the anastomosis can be performed between the transverse and the sigmoid colon. This will require mobilization of the right colon as well. Limited transverse colectomy may be considered for palliation. As mentioned, however, as a cancer operation it may be inadequate. Mobilization of the splenic flexure and left colon are explained in the section Rectal cancer surgery. The same principles are applicable here.

When the tumour invades the spleen or when the spleen is at risk, *en bloc* splenectomy may be indicated. In these circumstances the spleen and tail of the pancreas may need to be removed *en bloc* with the colon. Splenectomy in association with colonic resection for cancer is associated with a high morbidity and mortality rate. The incidence of unintentional splenic injury during mobilization of the splenic flexure is around 1%. Splenectomy increases the rate of postoperative sepsis in this group of patients.

Left hemicolectomy (partial colectomy)

Left hemicolectomy is performed for tumours involving the distal transverse colon, splenic flexure and descending colon. In this procedure, the proximal right branch of the middle colic and the inferior mesenteric artery should be preserved. The left colic artery is ligated. The anastomosis is performed between the mid-transverse and the upper sigmoid colon. The technique of the mobilization is described in the section Rectal cancer surgery.

Subtotal/total colectomy

Subtotal colectomy involves the removal of most of the colon with an anastomosis between the ileum and the sigmoid or descending colon. Total colectomy involves the removal of the entire colon with an ileorectal anastomosis at the rectosigmoid junction. It is a more extensive operation with the advantage of an easier to perform anastomosis and maximum harvesting of lymph nodes. Subtotal or total colectomy may be indicated for the following:

- synchronous left- and right-sided tumours
- multiple tumours (benign or malignant or both) are present
- FAP and IBD
- HNPCC
- when a resection has been previously performed
- technical limitations.

Complete mesocolic excision

Surgery for rectal cancer has become standardized with the implementation of TME. Until recently, there had not been any attempts to standardize surgery for colon cancer. Hohenberger *et al*. suggested the equivalent to the TME procedure, the complete mesocolic excision (CME) of the colon. The concept of CME is based on the same embryological principles of TME. It involves sharp dissection of the visceral plane from the retroperitoneal one with high ligation of the draining/supplying vessels, aiming to prevent any injury of the visceral fascia layer that may potentially result in spillage of tumour cells and dissemination of the disease.

For right colon cancers the mobilization of the duodenum with the pancreatic head (Kocher manoeuvre) and the mesenteric root up to the origin of the superior mesenteric artery is essential to maximize the exposure of the supplying vessels. The attachments of the mesenteric plane to the duodenum and the uncinate process are dissected in order to gain full access to the superior mesenteric vessels. The right colon is fully mobilized and therefore can be twisted (clockwise) to provide easier access to the central part of the superior mesenteric vessels. The ileocolic and, when present, the right colic vessels should be divided at their origin from the superior mesenteric vessels. For caecal and ascending colon cancer, ligation of the right branches of the middle colic vessels is adequate. The colon is divided at the level of the middle colic vessels. For hepatic flexure cancer, the transverse colon is resected at the splenic flexure.

For transverse colon cancers, the proximal part of the colon can be preserved and anastomosed to the sigmoid colon. The right colic vein drains into the superior gastroepiploic vein. It is usually ligated first to prevent severe haemorrhage from accidental injury. The superior mesenteric vessels are subsequently exposed, as the veins supplying the arteries are divided centrally. Attempt should be made to preserve the surrounding autonomous nervous plexus to avoid the risk of bowel dysfunction. If there are suspicious pancreatic lymph nodes they can be dissected off the pancreatic head with central ligation of the right gastroepiploic artery. The superior pancreaticoduodenal artery is usually preserved.

For left colon tumours, mobilization of the splenic flexure with the left colon is performed. This is described in more detail in the section Rectal cancer surgery. Once the left colon is completely dissected off the retroperitoneum, the greater omentum is divided from the transverse colon to facilitate full exposure of the lesser sac and the subsequent division of the final two layers of the transverse mesocolon at the lower edge of the pancreas. It is important that during this process the integrity of the mesocolon is maintained in a similar way to the mesorectal excision for rectal cancer. The root of the inferior mesenteric artery is usually preserved. The left colon artery is centrally ligated including the lymph nodes at the origin of the inferior mesenteric artery. Preservation of the superior hypogastric plexus is also important.

For cancers of the mid-descending colon down to the sigmoid colon, the root of the inferior mesenteric artery and the inferior mesenteric vein below the pancreas are ligated. The proximal colonic division is performed at the point between the left transverse colon and the proximal descending colon and depends on the tumour site. The resection is performed distally in the upper third of the rectum.

The implementation of CME can significantly increase lymph node harvesting. Hohenberg *et al.* harvested a median of 32 lymph nodes. The same study showed that in patients with N0 disease ($n = 314$) the 5 year survival was improved by almost 6% when the number of harvested lymph nodes was 28 or more (96.3% vs 90.7%). This difference was shown to be statistically significant ($p = 0.018$). The same conclusion was drawn for patients with positive lymph nodes ($n = 383$). There was an improvement in the 5 year survival by 6.1% when the number of harvested lymph nodes was 28 or more (71.7% vs 64.6%). This however did not reach statistical significance ($p = 0.088$). The cut-off of 28 lymph nodes was found to be an independent predictive factor in disease-free survival. At present, there are no clear guidelines about which colonic tumours are suitable for an open CME approach vs a laparoscopic colectomy. A 'low-risk' right colon cancer (T1, T2, T3a/b, node negative, extramural venous invasion (EMVI) negative) may be best managed with a standard laparoscopic right hemicolectomy. CME surgery may be reserved for 'high-risk', T3c/d, T4, node-positive, EMVI-positive colon cancers where a high nodal dissection may be indicated to maximize the oncological clearance of such cancer.

Wound closure

Abdominal wall closure depends on the surgeon's preference. In the UK a loop 0-PDS or 0-nylon suture is commonly used for mass closure using a continuous technique. Other sutures such as polyglyconate can be used. There is no evidence in the literature to suggest significant differences between the above sutures in terms of wound infection rates, dehiscence and hernia formation. The skin is closed either with clips or with a 3/4-0 monocryl subcuticular suture.

Rectal cancer surgery

Historical perspective

The first attempts to surgically treat rectal cancer took place during the nineteenth century, when Giovanni Morgagni was the first to suggest that resection of the rectum could be used as a treatment of rectal carcinoma. Operating at that time was performed without general anaesthesia and haemostasis, increasing significantly the rates of mortality and morbidity. The first two reported resections were unsuccessful and the patients did not survive the operation.

The first attempt to resect the rectum was performed in 1739 by Jean Faget, who intended to drain an ischioanal abscess but instead found a perforated rectal cancer. The patient died following the operation as he was left with an 'uncontrollable anus.' The second recorded attempt to resect an annular obstructing lesion also resulted in the patient not surviving the operation.

It was not until 1826 that the first successful resection of the rectum was performed by Lisfranc. This success was followed by five more successful operations from a total of nine that Lisfranc performed in the following years. These procedures were performed without anaesthesia and haemostasis, carrying significant risks of postoperative morbidity, with unbearable pain, sepsis and major bleeding being the commonest complications. During that era of surgery, the operation was considered successful if the patient was discharged home following the procedure. The patients who survived the operation would definitely return with recurrence and die.

The application of aseptic technique principles and the introduction of general anaesthesia enabled the performance of more advanced procedures, with Theodor Kocher proposing the transacral resection of the rectum, which was further developed by Paul Kraske, who aimed to improve the operative exposure. Carl Gussenbauer followed, using an abdominal approach to resect the tumour and Hochenegg used a 'pull-through' (duerchzug) technique to perform a rectoanal anastomosis.

Abdominoperineal resection

The first abdominoperineal resection was performed by Vincent Czerny in 1884, when the tumour could not be removed 'peritoneally' and the procedure had to be completed abdominally after turning the patient supine. Taking advantage of the larger operating field, Ernest Miles was the first to perform an 'anatomically correct' resection of the rectum removing the draining lymph nodes, at the same time combining the abdominal and perineal approach.

Using this technique, Miles reported a reduction in the recurrence rate to 29.5%. The most common complications were sepsis and blood loss resulting in a mortality rate of 31% as blood transfusion and antibiotics were not yet available. The patient's quality of life was significantly impaired, since the formation of colostomy was standard practice at the time, and the extensiveness of the procedure meant that the autonomic nerve plexuses were damaged and consequently sexual and urinary function would be adversely affected. Many surgeons suggested that the Miles procedure was too radical, carrying significant risks of genitourinary and bowel dysfunction with adverse psychosocial implications. This led Cuthbert Dukes (Dukes' classification of colorectal tumours) to further investigate the resected specimens following the Miles procedure and demonstrate that Miles overestimated the tumour spread within the regional lymph nodes and that the majority of lymph nodes were close to the level of the primary tumour. The modern abdominoperineal resection involves an extralevator dissection, which involves detaching the levator ani muscles from their insertion at the lateral pelvic sidewall. The posterior insertion of the levator ani muscles can be detached by removing the coccyx and extending the dissection proximally by lifting the presacral (Waldeyer's) fascia off the distal sacrum *en bloc* with the resected specimen. This has recently been popularized as the 'prone extralevator APER (abdominoperineal excision of rectum)', although the same oncological clearance can be obtained with the patient in the Trendelenburg position.

Anterior resection

As a result of the poor quality of life following surgery, surgeons had started to focus on less radical resections that led to the now widely established anterior resection. Donald Balfour described the technique of the anterior resection with the performance of an end-to-end anastomosis. Dukes' findings about Miles' overestimation of tumour spread led surgeons to investigate further the potential of anterior resection as an alternative for cure, with the advantage of maintaining bowel continuity. Anterior resection was not widely adopted until 1948, when Claude Dixon demonstrated that it was safe to perform with an improved mortality and 5 year survival rate of 2.6% and 64% respectively. Anterior resection was only performed for tumours that were more than 5 cm above the anal verge, as it was believed at the time that 5 cm distal margins were essential to achieve a curative resection. The possibilities for further development of anterior resection were realized after a study demonstrated that distal margins up to 2 cm do not compromise the oncological result in terms of survival and local control. Sphincter preservation procedures were the standard in cases where the sphincter was not directly involved by the tumour. In 1986, the creation of a colonic pouch combined with coloanal anastomosis was proposed, and improved bowel functional outcomes.

Total mesorectal excision

In the late twentieth century, Professor Heald popularized the technique of TME based on the embryological development of the hindgut, after studies demonstrated that resection margins of 2 cm should be considered as safe. Professor Heald initially suspected the value of the mesorectum in 1982, which led him to further investigate the 'holy plane' around the embryological hindgut and propose a standardized oncological rectal surgery by performing TME of the rectum, which could potentially improve the oncological outcomes while preserving the sphincter. In 1992, he successfully demonstrated that TME significantly reduced the local recurrence rate, especially for patients with Dukes' B and C stage cancers. These, along with the findings that a 2 cm distal surgical margins is sufficient, have probably been the most significant milestones in rectal cancer surgery, and led surgeons to widely adopt the technique of TME, welcoming the option to preserve the sphincter and bowel function without compromising the oncological outcome.

Extended lateral pelvic sidewall lymphadenectomy

Between 30% and 40% of patients treated for rectal cancer present with lymph node metastases. These can occur either along the mesorectal nodal chain along the inferior mesenteric artery nodes in around 40% of patients, and/or to the lateral pelvic lymph nodes (along the obturator, internal iliac and medial aspect of the external iliac artery) in 10–25% of patients. The presence of lateral lymph node metastasis in rectal cancer was first reported in the 1950s, and nodal involvement has subsequently been shown to be associated with a poorer prognosis, higher incidence of local recurrence and reduced survival.

In Western countries, TME has become the operation of choice for low rectal cancer. The technique is associated with a low incidence of local recurrence and facilitates sparing of the pelvic autonomic nerve plexuses. With the introduction of neoadjuvant radiotherapy combined with TME, local recurrence rates of less than 10% have been achieved. Western surgeons do not utilize extended lymphadenectomy regularly, and this may pose a risk of local recurrence in the pelvic sidewall in patients who have undergone potentially curative surgery without preoperative radiotherapy. Whether pelvic sidewall lymph nodes should be considered metastatic disease or part of the regional lymphatics is a contentious issue.

Based on the above, Japanese surgeons have adopted the technique of extended lymphadenectomy to supplement TME, aiming to minimize local recurrence and improve survival. The disadvantages of extended lymphadenectomy include increased rates of autonomic nerve dysfunction such as urinary and sexual impairment and intraoperative adverse events. There has also been interest in neoadjuvant radiotherapy in addition to TME and Nagawa *et al.* showed that the addition of extended lymphadenectomy to surgery, when neoadjuvant radiotherapy had been previously applied, did not significantly alter the local recurrence and survival rates.

Japanese surgeons proposed the lateral pelvic sidewall lymphadenectomy as an additional procedure to standard rectal cancer surgery, with a number of studies demonstrating controversial results. The introduction of neoadjuvant radiotherapy to the management of rectal cancer came late in the twentieth century, demonstrating reduced recurrence rates but without any significant impact on the patients' survival.

All the efforts of these great surgeons have significantly changed the management of patients with rectal cancer, improving oncological outcomes and quality of life. However, the recurrence rates following curative resection of the primary tumour are still considered high, necessitating more radical resection (exenterative procedures), such as abdominosacral resection of the rectum and total pelvic exenteration, which involves the removal of the adjacent pelvic organs and structures.

Survival and local recurrence

The main goal of surgery for rectal cancer is to achieve complete tumour removal with minimal complications and acceptable urinary, bowel and sexual function. The improvements in surgical technique and the awareness of the importance of adequate excision margins has enabled improvement of oncological outcomes by achieving higher rates of complete tumour removal (R0 resection), which is reflected in the reduction of locoregional failure rates. Although surgery and radiotherapy play a crucial role in the ultimate outcome, it appears that tumour pathology is the most important factor for overall and local recurrence-free survival.

Local excision for rectal cancer

The original criteria for local excision included small tumours (Dukes' A) of the rectum, which were mobile and accessible to endoanal removal. Morson *et al.* amplified the purely clinical criteria of Lockhart-Mummery to include two phases: the clinical assessment and the subsequent histopathological findings after removal. These included complete microscopic excision, tumour confined to the rectal wall, absence of lymphovascular invasion and a well-differentiated tumour. In patients having a local excision based on the clinical criteria of size under 3 cm, mobility and the 'feel' of being confined to the rectal wall, the final therapeutic decision would be determined by the pathology. If the above histopathological circumstances were not fulfilled, the patient was advised to undergo total anorectal excision.

The present

Local excision has been used for cure or for palliation. In the latter, it can be applicable in patients with disseminated disease or in those with comorbidity precluding major surgery or in the elderly, provided in both cases that the tumour itself is small enough and sufficiently localized to the rectal wall to allow local excision to be technically possible.

The rationale for local excision is that a tumour confined to the rectal wall has a low likelihood of having simultaneous lymph node metastasis. Only a few patients are amenable to local excision for cure and make up only about 5% of the total number presenting with rectal cancer. They must be selected very carefully.

Over the last 20 years, two schools have developed in regard to local excision. The first has reduced the indications of those of Morson to include only T1 or a few so-called favourable T2 tumours. The second has explored the use of adjuvant radiotherapy in combination with local excision for growths of T2 or even T3 stages.

Pathology

Local excision is based on the premise that a tumour of early stage, for example T1 or T2, will have a low prevalence of regional lymphadenopathy. Thus, local excision removing the growth in its entirety according to the subsequent pathological examination would be sufficient to cure the tumour. The prevalence of regional lymphadenopathy for each T stage is therefore one key to case selection and the appropriateness or not of local excision. The best data available are those derived from the Erlangen Tumour Centre and University of Surgical Clinic 1986–1990. Hermanek studied the incidence of regional lymphadenopathy for each pT stage in a large series of patients undergoing major resection of rectal cancer. Similar rates for T1 (10%) and T2 (17%) were reported by Blumberg *et al.* The incidence of regional lymph node metastases by depth of tumour penetration are shown in Table 30.16.

In the light of these data, it is clear that local excision would result in up to a fifth or more of tumours confined to the rectal wall (T1 and T2) being inadequately removed owing to the presence of regional lymphadenopathy. Even in the T1 category, where the tumour is sessile, the Haggitt staging system has shown regional lymphadenopathy to be present in 10% or less of Sm1 and Sm2 tumours but in as much as 25% for Sm3 stage.

The second pathological consideration is that of the morphology of the tumour. This has received little attention, but the difference in cancer-specific outcome after local excision of exophytic and non-exophytic carcinomas is significant. In a series of 91 patients followed for 5 or more years, local excision of an exophytic carcinoma (polypoid, 30 sessile non-ulcerated, 26) resulted in a cancer-specific 5-year survival rate of 100% compared with 71% for a non-exophytic carcinoma (ulcerated, 15; flat raised adenocarcinoma, 20). Exophytic morphology was more likely to be associated with an earlier T stage with 48 (86%) of the 56 exophytic and only nine (26%) of the 35 non-exophytic carcinomas being of stage T1. Three (5%) of the 56 in the former group and nine (26%) of the 35 in the latter were stage T3.

Based on these facts, local excision as the only therapeutic modality should be offered for T1 tumours with exophytic morphology. These attributes are most likely to be found in a mobile tumour of 3 cm or less in diameter with well-differentiated

Table 30.16 Incidence of regional lymph node metastases by depth of tumour penetration

pT stage	*n*	Lymph node metastases (%)
T1	102	4.9
T2 (inner muscularis propria)	201	17.9
T2 (outer muscularis propria)	261	26.8
T3 (slight <1 mm)	278	45.7
T3 (extensive)	505	69.9
T4 (other organ involvement)	153	71.9
T4 (serosal penetration)	88	85.2

histology. It is still essential to adopt the combined clinical and histopathological policy whereby any excised tumour with incomplete excision, deeper penetration of the rectal wall than T1, lymphovascular invasion or poor differentiation should be discussed with the patient regarding subsequent major surgery.

Case selection

Based on the above criteria, the means of case selection include digital examination, endorectal ultrasound and MRI. Digital examination should not be underestimated. It is clear that it is not as accurate as ultrasound, being subjective and relatively insensitive. Nevertheless, it is only through digital examination that the possibility of local excision is considered in the first place, and the method can identify the T1 or T2 tumour with an accuracy of around 70%. Only those tumours identified as being small and mobile will be referred for ultrasound with local excision in mind.

Ultrasound has an accuracy of over 90% in assessing penetration of the rectal wall (the interface between T2 and T3). In a study of 351 patients, its sensitivity and specificity were reported for all T stages. Ultrasound correctly staged 49 of 71 T1, 104 of 149 T2 and 92 of 131 T3 tumours. Overstaging in the T1 group occurred in 29% and in 13% for T2 tumours. In a series of 356 patients, sensitivity and specificity for stage T1 were 86% and 94% respectively. Thus ultrasound was able to identify 69–84% of T1 cases in these studies. MRI is inadequate for identifying T1 and T2 tumours. It has a sensitivity of 40–60% and specificity of just under 90%.

Nodal involvement is more difficult to determine preoperatively. Indeed, there is currently no reliable means of doing so. Thus the overall accuracy reported by Garcia Aguilar *et al.* was 68% for non-involved and 52% for involved nodes. Mackay *et al.* obtained sensitivity and specificity values of 67% and 90%. MRI may, however, be able to improve on these figures. The finding on MRI of an irregular border or mixed signal intensity within a given node was associated with nodal involvement by tumour in 92% and 91% of nodes showing these features compared with 6% and 8% of nodes when they were absent. Further data on this are required.

Summary

A combination of digital examination, endorectal ultrasound and MRI will be able to select the T1 tumour in 70% or more of patients with this T stage. The added features of small size, exophytic morphology, accessibility and absence of poor differentiation all contribute to identifying the rectal carcinoma suitable for local excision or transanal endoscopic microsurgery (TEMS).

Technique

Endorectal local excision or TEMS are now the preferred methods. In the past, some surgeons used the posterior approach (Kraske, or that modified by York Mason), but this has the disadvantage of opening up the anatomical envelope of the rectum, which is oncologically unsound and will also make any subsequent major surgery more difficult than if an endorectal removal had been carried out.

For low rectal lesions (within 3–4 cm from the anal verge) an anal retractor and good lighting (a head light is recommended) are necessary; a full-thickness disc of rectal wall including the growth and a 1 cm margin of normal rectum are removed. Lesions in the mid-rectum are more accurately removed using a TEMS approach or an anal SILS port (single incision laparoscopic surgery). The former technique remains the gold standard for local excision, as the equipment provides three-dimensional endoscopic vision and the specialized angled instruments facilitate accurate removal of the rectal lesion and closure of the rectal defect.

Histopathology

The specimen is then pinned out onto a cork lamina and placed upside down in a formalin pot to fix. The pathologist will thus be able to orient it and take sections from the deep and circumferential margins. The report will state the following: completeness of excision, the presence or absence of lymphovascular invasion, the degree of penetration of the rectal wall and the histological grade.

An incomplete or doubtfully complete removal will make resection advisable. It is reasonable to consider 2 mm of clear margin as adequate. The risk of lymph node involvement is mainly determined by T stage, the histological grade of differentiation and vascular and lymphatic invasion. When invasion is confined to the submucosa, lymph node metastasis is present in only 5% of cases, but this incidence rises to over 20% in T2 tumours (see above).

Poor differentiation has a high risk of lymph node metastasis. Hermanek found nodal involvement in 77% of poorly differentiated cancers. Nevertheless, high-grade tumours are particularly uncommon in early rectal cancer with an incidence of less than 5%. Lymphovascular invasion has been usually considered by most pathologists as an unfavourable feature, with an increased risk of lymph node metastasis. Submucosal lymphatics are often difficult to see and there is wide interobserver variation. In a series of 81 malignant polyps, the assessment of lymphatic invasion may sometimes be too subjective to draw any valid conclusion. In the same series, submucosal venous invasion was found to have no prognostic importance.

Where any of the above pathological features indicate likely failure of clearance locally or a high chance of lymphadenopathy to be present, the patient must be advised accordingly. The risk of local recurrence and lymph nodes metastasis must be balanced by the morbidity and mortality of surgery. The decision will depend on the patient's wishes in the light of discussion about the pathological and clinical prognostic factors. Discussion may then result in agreement whether or not to undertake a total anorectal excision.

Results

Local excision is an uncommon treatment, and most reports in the literature contain small numbers of patients often accrued over many years. Thus case selection in many instances will have been made before the introduction of endorectal ultrasound and more recent imaging.

Several useful series have been reported. In a large series from the Mayo Clinic of 234 patients including 93 with a T1 and 141 with a T2 tumour, local failure occurred in 19%.

Killingback reviewed 63 cases of local excision out of a total of 493 patients with rectal cancer treated between 1969 and 1984. After excluding 25 with a malignant polyp, nine of the remainder developed local recurrence and seven (18%) died. Of the nine failures only five were suitable for a salvage operation, of whom three died. Whiteway *et al.* in a report about local excision of an ulcerating carcinoma in 27 patients, found a cancer-specific mortality of zero when the policy of subsequent major surgery following pathological examination of the specimen had been fulfilled. Of 26 patients who survived the operation, two refused further surgery despite being advised to have it. Both died of cancer. In the remaining 24 patients, five underwent subsequent major surgery, two of whom died of cancer. Therefore, the overall failure due to cancer occurred in 8% (two patients). In a report of 110 patients treated at St Mark's Hospital from 1948 to 1973 by local excision, there were only six failures. This series was updated to 1984. Of 167 patients with reliable follow-up of over 5 years, 152 were available for analysis. The results emphasize the importance of histological grade. Thus, local failure occurred in only one (2%) of 56 well-differentiated tumours compared with 12 (19%) of 64 patients with a moderately differentiated tumour.

Using more limited criteria for case selection Hermanek and Gall reported on 249 patients with an early carcinoma of whom 53 patients underwent local excision. The 5-year survival rate in this group was 95%. Of 130 patients having a major resection for a tumour of the same T stage, the 5-year survival rate was 100%.

The question of whether major resection or local excision should be performed was considered further by Mellgren *et al.* They reported the results of local excision in 108 patients treated for a T1 or T2 rectal carcinoma. These were compared with 153 patients with T1N0 and T2N0 lesions treated with abdominal radical surgery. No patients had radiotherapy. Local recurrence after local excision occurred in 18% of T1 and 47% of T2 tumours. This rate was compared with zero for T1 and 6% for T2 tumours after radical surgery. The authors pointed out, however, that survival was little different between the two groups when the results of salvage surgery after failed local excision in line with the policy of Morson were taken into account.

A more recent comparison of outcome of T1 tumours treated by local excision or radical surgery was reported by the Memorial Hospital New York. Of a total of 319 patients with a T1 tumour treated over a 17-year period, 151 underwent local excision and 168 a radical resection. The tumour was located in the rectum, significantly higher in the latter group by 2 cm and the diameter was on average 4 mm greater. Follow-up was 48 (1–145) and 58 (2–192) months respectively. The actuarially estimated 5-year local recurrence rates were 15% and 3% respectively ($p < 0.01$) and cancer-specific survival rates were 93% and 97% ($p < 0.05$). The level of the tumour may have been a factor in these differences, but lymph node metastasis may also have been important since the incidence of nodal involvement in the radical surgery group was 18%.

The evidence therefore indicates that radical surgery is less likely than local excision to be followed by local treatment failure and longer survival. A similar conclusion was drawn by Naseimbeni *et al.* regarding local recurrence and overall survival. These data emphasize the degree of risk and the importance of case selection associated with local excision.

Local excision combined with adjuvant therapy

While one line of development of local excision has been to reduce the indications of local excision, some authors have combined it with adjuvant postoperative radiotherapy with or without chemotherapy in an attempt to expand them. In reviewing the results, it should be borne in mind that major surgery for T2 and early T3 tumours has a low rate of local recurrence.

With the exception of Bleday *et al.* who reported no cases of local recurrence among 21 T2 tumours at a follow-up of 40 months [although local recurrence occurred in two (25%) of eight T3 cases], the experience of others has been less satisfactory. Local recurrence after local excision of T2 tumours in patients followed from 41 to 73 months ranged from 14% to 24% out of a total of 162 patients. For 37 patients with a T3 tumour, the rate of local recurrence ranged from 23% to 40%.

It is difficult to interpret the results in some incidences since these are not controlled data, and exclusions were made in certain instances. For example, Steele *et al.* in reporting 161 patients with a T1 or T2 tumour undergoing local excision followed by postoperative chemoradiotherapy, 51 were excluded from analysis owing to a positive margin, leaving eight patients with local recurrence among the remaining 110 patients. These data are summarized in a review by Sengupta and Tjandra.

It should be remembered that, for tumours of these T stages (including the early T3 case), major surgery is likely to be curative and also that cure can be achieved by ultra-low anterior resection where the growth is in the lower third involving where necessary an intersphincteric dissection. The decision for local excision is therefore a matter of great responsibility for the clinician.

Transanal endoscopic microsurgery

TEMS uses a sophisticated rigid endoscopic system having binocular optics, suction and carbon dioxide insufflation with sealable ports on the instrument to enable the introduction of specially designed instruments. The sigmoidoscope is fixed to the operating table, which can be adjusted. Formal training is required to perform this procedure (Figure 30.49). The operating field views are excellent and all the instruments are available in right- and left-handed versions. A high-frequency electrode coagulator is used for cutting.

The technique is similar to conventional endoanal excision. The margin of the tumour is marked using coagulation at a distance of 10 mm from the edge of the tumour circumferentially. The incision is then made with the electrocutter, and haemostasis is achieved using coagulation diathermy as necessary. The tumour is removed with a full thickness of rectal wall. The subsequent defect is sutured together endoscopically. The sutures used in this procedure use a short length of thread and are fixed by silver clips rather than knots.

TEMS vs local excision for early low rectal cancers

Although TEMS is considered a better surgical technique than local excision, it cannot be used in ultra-low rectal cancers.

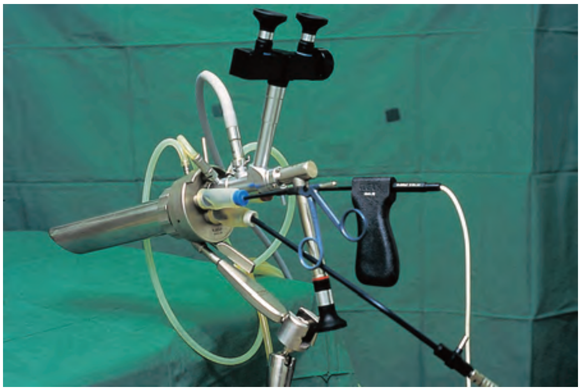


Figure 30.49 Buess operating protoscope for transanal endoscopic microsurgery.

Local excision may be a better option for this type of tumour. A recent study showed that the resection margins were more often positive in the local resection group (16%) than in the TEMS group (2%; $p = 0.017$) without any significant differences in terms of postoperative complications between the two groups. The authors of that study interpreted these results as to be related to the better quality of resection that was facilitated by TEMS. However, these results might be partly due to case selection bias, as lower risk tumours of the upper rectum were included in the study.

Two randomized controlled studies in Europe are currently investigating the differences of TEMS vs TME surgery, the results of which are not available. The reported local recurrence rates in the literature following TEMS and local resection are variable. Studies have reported local recurrence rates for TEMS ranging up to 13% for patients with T1 tumours and up to 80% for patients with T2 tumours. Similar rates have been reported for local excision, with local recurrence rates reaching up to 40% for patients with T1 tumours and up to 60% for patients with T2 tumours. TEMS may provide better views especially for higher rectal tumours, whereas local excision may be more suitable for lower rectal cancers

Total mesorectal excision

TME is currently considered the gold standard for the surgical management of low and mid-rectal cancer. It is still debatable whether a TME should be performed for upper rectal cancers, as there is not adequate evidence to support it. When the tumour involves or is close to the anal sphincter complex, abdominoperineal resection is indicated. The performance of proper TME was demonstrated by several independent groups to reduce local recurrence rates. This initiated the organization of master classes globally to teach colorectal surgeons the technique.

Technique

The patient is positioned in the extended Lloyd-Davis position. A urinary catheter is advised, as it drains the bladder and therefore reduces its size and consequently increases the operating field. It can also be used to monitor the urine output during the procedure.

A midline incision from the symphysis pubis towards the xiphisternum is made. In selected patients a lower midline incision can be performed instead. When the peritoneal cavity is accessed, it is important to ensure that the disease is still operable by looking for liver metastases, intra-abdominal dissemination, distant lymphadenopathy and fixation to nearby structures. The small bowel is then packed away into the right upper quadrant to optimize the operating field. A self-retaining retractor is used to keep it in place.

Mobilization of the descending (left) colon is traditionally performed from lateral to medial. The congenital adhesions between the sigmoid colon on the left lateral wall are divided with the use of coagulation diathermy exposing the peritoneal reflection. The peritoneal reflection is then incised to reveal the embryonic plane of dissection (white line), which must now be developed using countertraction and coagulation diathermy. Haemostasis is ensured with the use of monopolar or bipolar diathermy. When the sigmoid mesocolon is reached, the hypogastric nerves will be in view. The left ureter is identified and swept laterally to prevent any injury. The ureter is a whitish, non-pulsatile cord and shows peristaltic activity when gently pinched with forceps. The plane anterior to the hypogastric nerves is followed down into the pelvis until the plane between the mesorectum and the sacral fascia has been identified. The rest of the left colon is mobilized by extending the plane of dissection superiorly up to the splenic flexure.

Any obvious adhesions between the omentum and the splenic capsule are divided to prevent haemorrhage from tearing the splenic capsule when traction is placed on the omentum. This is only performed when the spleen is visible. Diathermy is used to divide the peritoneal connections between the lateral abdominal wall and splenic flexure. The plane between the omentum and the transverse mesocolon is identified and entered, by applying traction on the transverse colon and the attached greater omentum. This plane is developed and followed to the level of the splenic flexure providing adequate and quick mobilization of the splenic flexure.

With the division of the splenic flexure completed, the dissection of the peritoneum continues until the level of the aorta, where the origin of the inferior mesenteric artery can be found. It is dissected, ligated and divided. The inferior mesenteric vein is found superiorly at the inferior border of the pancreas and is also dissected and ligated close to its origin. This will complete the mobilization of the large bowel from the transverse colon to the distal sigmoid. Good perfusion is maintained up to the top of the rectum. This degree of mobilization will ensure that the anastomosis is tension free.

The colon is usually divided at the apex of the sigmoid colon (for low rectal cancers) unless this is significantly affected by diverticulosis, when a higher division at the descending colon may be more appropriate. The sigmoid mesentery is divided up to the colonic wall, including the ligated trunks of both the inferior mesenteric artery and vein. The proximal bowel is packed up into the left upper quadrant of the abdomen.

The application of forward traction to the rectum will reveal the plane between the mesorectum and sacral fascia with the hypogastric nerves lying on it. This is gradually developed using the coagulation diathermy, ensuring that the nerves are not injured and separated from the mesorectal surface. Preserving the nerves reduces the risk of damaging the parasympathetic nerves on the lower lateral wall of the pelvis and therefore the risk of impairing the bowel, sexual and urinary function. Dissection is then performed to the lateral sites of the rectum with the use of a curved St Mark's retractor and gentle countertraction of the rectum. When the planes are adequately dissected, anterior dissection commences.

This part is slightly different between the two genders. In females, the initial plane of dissection lies between the uterus and the mesorectum. Therefore, it may be useful to suture the uterus to the anterior abdominal wall. The division of the peritoneum is performed at the plane above and anterior to the apex of the rectovesical (males) or the rectouterine pouch (females). The anterior wall of the rectum is pulled posteriorly. The St Mark's retractor is positioned anteriorly to the rectum. In males, the plane continues between the anterior mesorectum and the seminal vesicles and in females between the anterior mesorectum and the vagina. With the completion of the anterior dissection, the posterior and lateral aspects are reviewed to ensure adequacy of the whole dissection. The lateral ligaments are not clamped, as this is likely to damage the parasympathetic plexus. The nerves from this plexus that supply the rectum are divided.

The pelvic floor should be reached. Lloyd-Davies scissors or vessel-sealing diathermy device can be used to divide any posterior strands. When the dissection is completed the lowest part of the rectum is viewed as a muscle tube. A linear stapler is positioned into the pelvis and the lowest part of the rectum is carefully placed into its jaws. This should be positioned low enough to achieve complete tumour resection, but high enough to allow a further line of staples to be placed below it. The rectal stump is then washed out with povidone iodine solution. The hypothesis is that it may kill any intraluminal tumour cells but the evidence is currently controversial.

The circular stapler is inserted through the anal canal and the centre rod is advanced through the centre of the staple line. It needs to be ensured that the colon is not twisted and that the head of the gun is attached to the centre rod. The staple gun is closed and fired. The anastomosis can be checked by direct palpation. A rigid sigmoidoscope is also used to insufflate air in the lower part of the rectum while the anastomosis is immersed in warm normal saline to ensure that there are no air leaks. A leak can be sealed with an extra suture and retested. The resected donuts are also checked and sent for histopathology examination to test for tumour at the margins. The anastomosis can also be hand-sewn. In the NHS, this is dictated by the surgeon's preference. Where cost is an issue, hand-sewn anastomosis is more cost-effective but increases the intraoperative time. There is no evidence to suggest a significant benefit in terms of anastomotic leak or recurrence rate.

In low rectal anastomosis a temporary loop ileostomy may be performed to protect the anastomosis until it is healed. Gastrografin enema is performed 6–8 weeks following the operation to assess the integrity of the anastomosis. Reversal of the anastomosis is performed 6–8 weeks following the primary operation or the completion of adjuvant chemotherapy. In addition some surgeons may perform a colopouch, aiming to improve bowel functione outcomes.

Sphincter preservation ultra-low anterior resection

In very low rectal cancers, an ultra-low anterior resection can be performed and therefore preserve the sphincter. A coloanal anastomosis can be performed to maintain bowel continuity. A temporary stoma is necessary to protect the anastomosis. It can be reversed at a later stage following complete recovery from surgery and completion of any adjuvant medical therapy that may have been given.

Technique

Following completion of TME as described above, the transanal part of the procedure can be performed. A hooked self-retaining retractor is used for better access and vision. A mixture of epinephrine (adrenaline) (1:100 000) is injected initially into the submucosal plane and higher up into the intersphincteric plane. An initial circumferential mucosectomy is performed from the dentate line. The distal part of the internal sphincter can be maintained depending on the tumour's proximity. It is important to completely resect the mucosa, as incomplete resection will compromise the anastomosis.

With the completion of mucosectomy the lower part of the rectum can be resected. The distal part of the remaining sigmoid or descending colon is brought to the anus, on top of the internal sphincter where an anastomosis will be performed. A tension-free anastomosis is essential and therefore complete mobilization of the splenic flexure is important. In order to protect the anastomosis, a defunctioning ileostomy or colostomy is performed. This can be reversed at a later stage following a water-soluble enema that shows no anastomotic leak.

Published results suggest that partial or complete excision of the IAS with a hand-sewn anastomosis can be performed with a low operative mortality (1.6%), and acceptable rates of anastomotic leakage (<10%) and local recurrence (9.5%). The potential disadvantage of performing a restorative operation for very low rectal cancers, particularly when all or part of the IAS is removed, is the unacceptable functional outcomes following the anastomosis. The extent of internal sphincter excision has been shown to influence the postoperative resting tone. Several strategies can be employed to improve the functional outcome by increasing the volume of the neorectal reservoir (such as the colonic J-pouch).

Extralevator abdominoperineal excision of rectum

Abdominoperineal excision of the rectum is performed for very low rectal cancers or locally advanced primary cancers that may be invading the ischiorectal fossa or posterior vaginal wall. The procedure is divided into two parts, the abdominal and perineal part. Following completion of the abdominal part, the surgeon will start the perineal dissection. The abdominal part is usually performed first as it will enable the assessment for distant metastases. Dissemination of the disease may result in the procedure being abandoned.

The perineal part of the procedure follows the completion of the abdominal part, which is similar to TME as described above (except when the tumour is locally advanced – extensive resection may be required). The anus is sutured closed with the use of a pearl suture. An oval incision around the anus is made. Its size depends on the tumour size and position.

Exposure is essential for the perineal part. The St Mark's retractor is commonly used to increase the view of the operative field. The dissection is progressed to the ischiorectal fossa on each side and then posteriorly towards the coccyx. Although the coccyx can be preserved unless the tumour extends posteriorly, the senior author's preference is to routinely excise this *en bloc* with the rectum as it reveals the posterior surface of Waldeyer's fascia, which is then excised together with the levator ani muscles. Co-operation between the abdominal and perineal surgeon is fundamental at this point as the abdominal surgeon guides the perineal surgeon as to where the dissection needs to be performed.

The abdomen is entered using scissors after the perineal surgeon identifies the lateral border of the levators with his index finger. The muscle is subsequently freed laterally from its pelvic attachments. Bilateral resection of the levators will free the rectum posteriorly and laterally. This will include both levators with the specimen (extralevator APER), minimizing the risk of incomplete resection and the risk of local recurrence. The prone position has been popularized in Europe for the perineal aspect of the dissection; however, the same procedure can be achieved by the modified Lloyd-Davis position.

Lateral pelvic sidewall lymphadenectomy (extended lymphadenectomy)

Technique

The current lateral lymph node dissection is performed the same way as Professor Moriya described it in 1989. Following dissection and mobilization of the sigmoid and descending colon, the pelvis is exposed enabling the initiation of the lateral dissection. The retroperitoneum is dissected at the level of the bifurcation of the aorta, exposing the common iliac vessels.

It is performed either *en bloc* with the internal iliac vessels or by preserving them, depending on the presence of tumour close to the internal iliac vessels or on the evidence of solitary lymph node metastases respectively. The lateral pelvic sidewall lymphadenectomy starts with the separation of the internal iliac vessels from the lateral ligament medially. All the branches of the internal iliac vessels are identified, ligated and divided as close to their root from the internal iliac vessels as possible.

For cases in which the tumour is close to the iliac vessels, the lateral pelvic sidewall lymphadenectomy is performed *en bloc* with the resected specimen, including the unilateral internal iliac vessels with preservation of the superior vesical artery and the obturator nerve. When the lateral pelvic sidewall lymphadenectomy is performed to remove lateral lymph nodes that are suspicious for malignancy, the nerve-preserving approach can be utilized. The specimen is resected *en bloc* with the tumour while the lateral pelvic sidewall spaces are opened between the lateral aspect of the internal iliac vessels and the pelvic sidewall, exposing the lateral lymphatic tissue and enabling harvesting of the lymph nodes. Using the nerve-preserving technique, the obturator nerve and internal iliac vessels with their branches are preserved without disturbing the sacral nerve plexuses.

With the completion of the lateral pelvic sidewall dissection and the abdominal mobilization of the specimen and depending on the tumour's extent, a TME or other more advanced operation can be performed (i.e. pelvic exenteration, posterior exenteration, abdominoperineal excision of the rectum or an abdominosacral resection).

Results

A meta-analysis compared the value of extended and non-extended lymphadenectomy of the lateral pelvic wall nodes in patients with primary rectal cancer and showed no overall difference in the oncological outcome. There was a marginal benefit for the extended lymphadenectomy group in survival and local recurrence, in patients with a Dukes' C but not Dukes' B carcinoma. However, the procedure increased the intraoperative time and blood loss by 76.7 minutes and 536.5 mL respectively. It was also associated with higher incidence of urinary and sexual dysfunction.

Management of advanced rectal cancer

When cancer has progressed to surrounding tissues and anatomical structures, the surgery required to achieve complete clearance is more complex with a higher risk of incomplete resection, morbidity and mortality. This group of patients should be managed at tertiary centres where expert specialists such as oncologists, radiologists, surgeons and histopathologists are present and are supported by appropriate infrastructure.

In an effort to establish criteria that would enable better prediction of tumour resectability and outcome, a number of classifications have been proposed. These are shown in Table 30.17.

All these classifications aim to describe to what tissue the tumour is fixed, and hence establish the potential for and extent of resection in order to define selection criteria for surgery along with prognostic information. The Royal Marsden classification has been found by the authors to be the most useful, enabling better communication and understanding between radiologists and clinicians. It helps to select patients for surgery and surgical planning and at the same time provides detailed information on the extent of the disease and signifies prognosis.

Selection criteria for surgery

The decision for surgery is made after extensive discussions at the local multidisciplinary meeting and relies heavily on the findings of the available diagnostic modalities. It is the radiological findings that will determine tumour resectability. Therefore, accurate preoperative staging can help to establish the extent of local disease and the presence or absence of distant metastases, which is imperative when considering patients for exenterative pelvic surgery for recurrent colorectal cancer.

Distant metastases

The presence of distant metastases is normally considered as a contraindication for exenterative surgery for locally advanced primary rectal cancer. It is generally contraindicated

Table 30.17 Staging classifications for locally advanced primary rectal cancer

Study group	Classificatio	Definition		Outcomes
Mayo Clinic	Symptoms	S0	Asymptomatic	Pain-free patients had better survival
		S1	Symptomatic without pain	
		S2	Symptomatic with pain	
	Degree and site of fixation	F0	No fixation	More points of fixation resulted in more complications and worse survival
		F1	Fixation to one point	
		F2	Fixation to two points	
		F3	Fixation to >2 points	
Yamada	Pattern of pelvic fixation	Localized	Invasion to adjacent pelvic organs or tissue	5-year survival of 38%
		Sacral invasive	Invasion to lower sacrum (S3, S4, S5), coccyx, periosteum	5-year survival of 10%
		Lateral invasive	Invasion to sciatic nerve, greater sciatic foramen, lateral pelvic wall, upper sacrum (S1, S2)	5-year survival of 0%
Wanebo	Five stages	TR1	Limited invasion of the muscularis	
		TR2	Full-thickness penetration of muscularis propria	
		TR3	Anastomotic recurrences with full-thickness penetration beyond the bowel wall and into the perirectal soft tissue	
		TR4	Invasion into adjacent organs without fixation	
		TR5	Invasion of the bony ligamentous pelvis including sacrum, low pelvic/sidewalls, or sacrotuberous/ischial ligaments	
MSK	Anatomical region involved	Axial	Anastomotic, mesorectal, perirectal soft tissue, perineum (APER)	Axial-only recurrence has 90% likelihood of R0; lateral recurrence is associated with 36% likelihood of R0
		Anterior	Genitourinary tract	
		Posterior	Sacrum and presacral fascia	
		Lateral	Soft tissues of the pelvic sidewall and lateral bony pelvis	
Royal Marsden	MRI; planes of dissection	Central	Rectum or neo-rectum, intraluminal recurrence, perirectal fat or mesorectum, extraluminal recurrence	MRI diagnosis of tumour invasion within the lateral, posterior or in more than two compartments was associated with reduced disease-free survival
		PR	Rectovesical pouch or rectouterine pouch of Douglas	
		AA PR	Ureters and iliac vessels above the peritoneal reflection, sigmoid colon, small bowel and lateral sidewall fascia	
		AB PR	Genitourinary system	
		Lateral	Ureters, external and internal iliac vessels, lateral pelvic lymph nodes, sciatic nerve, sciatic notch, S1 and S2 nerve roots, piriformis or obturator internus muscle	
		Posterior	Coccyx, presacral fascia, retrosacral space, sacrum up to the upper level of S1	
		Inferior	Levator ani muscles, external sphincter complex, perineal scar (APER), ischioanal fossa	

AA, anterior above; AB, anterior below; APER, abdominoperineal excision of the rectum; PR, peritoneal reflection

due to the significant morbidity that is associated with this type of procedure. It has been demonstrated by some centres that synchronous or staged resection of metastases can have acceptable results in highly selected patients.

Resectable cancer

In the absence of distant disease, surgical resection of the advanced primary cancer is the only potentially curative option. Surgery for advanced primary rectal cancer includes a range of different procedures that depend on the extent of the disease and the specific organs/structures that are involved. Surgery is performed *en bloc* and is considered curative when the resection margins are free of microscopic disease (R0 resection). Microscopic or macroscopic residual disease at the resection margins are defined as R1 and R2 resection respectively.

Contraindications for surgical resection

It is essential to assess the patient's fitness for surgery before discussing surgery with the patient, since lack of fitness could be a contraindication when undergoing such a major procedure.

Surgery is contraindicated in the presence of:

- circumferential or extensive lateral pelvic sidewall involvement
- involvement of the iliac vessels
- bilateral ureteric obstruction
- sciatic nerve involvement
- para-aortic lymph node metastases
- irresectable distant metastases.

Limited tumour invasion to the lateral pelvic wall and invasion of the sacrum above the S2 vertebrae are considered relative contraindications since there are surgical options in both cases, but the likelihood of a complete resection is considerably lower.

Irresectable local cancer

Surgical resection and chemoradiotherapy can be used for palliation alleviating the patient's symptoms that are related to the organs/structure which have been invaded by tumour. It has been suggested that palliative resection can improve quality of life and pain relief. However, it is usually unsuitable because of the high morbidity rate related to this type of surgery. It is therefore important that patients are carefully selected for palliative procedures, taking into consideration possible comorbidities and their social circumstances, as the benefits from these procedures are short term.

Surgery for metastatic disease

It is now recognized that a proportion of patients with hepatic metastases from colonic cancer are suitable for liver resection. Approximately 10% of all patients with colorectal cancer are found to have synchronous liver metastases and half of these (5% of the total) have resectable lesions. A further 10% of patients will develop metachronous liver metastases within about 2–3 years of resection of the primary colorectal cancer. The important criteria for resectability of hepatic metastases from large bowel cancer are as follows.

- The primary tumour must be controlled with no evidence of extrahepatic metastases.
- All of the hepatic metastases must be resectable with an adequate margin of at least 2 cm of normal liver tissue.
- The operation must be performed with an acceptable mortality and morbidity.

Although the best results are associated with solitary metastatic lesions, it is possible to carry out extensive liver resections for multiple disease. Analysis of several large series indicates that multiple factors influence survival after hepatic resection. These are as follows.

- *Stage of the primary tumour.* The more advanced the stage of the primary lesion the worse the prognosis after liver resection.
- *Number of metastases and their location within the liver.* Patients with more than three metastatic lesions and those with bilobar disease tend to fare badly.
- *The interval of time between resection of the primary lesion and the appearance of the hepatic metastases.* The longer the period the better the prognosis.
- *Sex of the patient.* In general, women live longer than men after resection of liver metastases.
- *The presence of extrahepatic metastases.* This adversely influences the results of liver resection.
- *The operative mortality.* Modern surgical standards dictate that this should be under 5% even for a major lobectomy or extended right hepatectomy.

Although there has been no randomized trial there is little doubt that resection of liver metastases confers a survival advantage. With careful patient selection, hepatectomy for colorectal metastases is associated with a 5 year survival around 30%, which is the same prognosis as is associated with resection of a stage C cancer without evidence of distant metastases.

In patients with liver metastases which are not suitable for liver resection, then *in situ* ablation using cryotherapy or radiofrequency probes may be of value. Currently this is performed under ultrasound control, either at open surgery or laparoscopically. It seems to be associated with a modest prolongation of survival. Future developments include percutaneous *in situ* ablation using MRI-guided probes in an open magnet and the use of extracorporeal high-energy focused ultrasound to destroy tumour deposits (see also Chapter 24).

Laparoscopic vs open surgery

Laparoscopic colorectal surgery has evolved immensely during the recent years. Colorectal surgeon training is more specialized and there is focus in the use of laparoscopic surgery. There has been extensive interest in the application of laparoscopic surgery in colorectal cancer, with a significant number of publications investigating the short- and long-term benefits. The mean operative time is significantly longer when laparoscopic surgery is performed. Laparoscopic surgery is associated with less pain (and analgesia requirements), faster recovery of intestinal motility and return to normal diet, and significantly lower overall morbidity. These factors contribute to a shorter hospital stay. With the introduction of enhanced recovery protocols, however, these effects may be less significant.

Worldwide there is now a substantial published experience of laparoscopic surgery for colonic cancer, and the early concerns regarding laparoscopic surgery for colonic cancer have been dismissed by the results of several randomized controlled trials which have confirmed identical oncological outcomes and survival rates together with benefits in terms of reduced blood loss, less postoperative pain and faster recovery of gastrointestinal function.

These randomized controlled trials, meta-analyses and systemic reviews have confirmed no difference in the reported disease-free survival between laparoscopic and open surgery. In trials reporting on tumour recurrence, two randomized controlled trials reported zero event rates in both surgery groups. Worldwide, there have been only three reported instances of port site deposits. Four studies were included in a meta-analysis on hazard ratios for tumour recurrence in laparoscopic colorectal cancer surgery. No significant difference in recurrence rate was observed between laparoscopic and open surgery. Likewise, no significant difference in tumour recurrence between laparoscopic and open surgery for colon cancer was observed between the two groups. In a Cochrane review, 33 randomized clinical trials comparing laparoscopically assisted vs open surgery for colorectal cancer were identified. Twelve of these randomized controlled trials involving 3346 patients reported long-term outcome and were included in the analysis. No significant differences in the occurrence of incisional hernia, reoperations for incisional hernia or reoperations for adhesions were found between laparoscopically assisted and open surgery. Local recurrence rates at the site of the primary tumour were similar (5.2% vs 5.6%). The review identified four randomized controlled trials for rectal cancer (714 patients) and found similar local recurrence rates (7.2% vs 7.7%). There were no differences in the occurrence of port site/wound recurrences. This review also reported similar cancer-related mortality for colon cancer

(1575 patients, 14.6% vs 16.4%) and for rectal cancer (578 patients, 9.2% vs 10.0%).

In the UK, NICE now supports and advocates the laparoscopic approach, provided the surgeon is fully trained. The decision about whether to use open or laparoscopic surgery should be made after informed discussion between the patient and the surgeon, taking into consideration the patient's condition and suitability for laparoscopic surgery, the risks and benefits of the two approaches, and the surgeon's experience in laparoscopic colorectal surgery. Although reported for all ages, the benefits of laparoscopic colorectal cancer are especially marked in elderly patients.

Postoperative outcomes

Approximately 40% of patients with colorectal cancer will die from the disease within 5 years of surgery. Eighty per cent of recurrences are seen during the first 2 years after primary tumour resection, thus there has been discussion to suggest the intensity of the follow-up regimen should be highest in the first few years.

The aims of follow-up for colorectal cancer include:

- management of postoperative complications
- early detection of recurrence, either local or distant
- early detection of new primary tumour
- aids decision-making for possible adjuvant therapies
- audit of outcomes
- assess quality of life
- patient reassurance.

Most studies that have compared aggressive follow-up with less aggressive strategies have suggested that increased surveillance has a positive trend with survival advantage. This may be due to earlier discovery of treatable recurrences amenable to curative reresection. However, there is meta-analytical work that has suggested that although intense follow-up may pick up some asymptomatic recurrences, the numbers needed to treat are very high and this leads to high labour intensity and poor cost effectiveness. The same work suggested that the focus of colorectal cancer follow-up should shift from the early detection of recurrence towards quality of life assessment and patient support.

Patient follow-up involves clinical review, serum tumour markers, endoscopy and radiology.

Clinical review

Several randomized studies have shown possible benefits but because of relatively small sample sizes the results were not statistically significant. However, a meta-analysis showed that more intensive follow-up leads to an increase in 5 year survival advantage. A previous meta-analysis was unable to show any significant increase in 5 year survival rate for intensive follow-up of patients with colorectal cancer operated on for cure, but this study did not include the use of CEA for early detection of recurrence and also included data from two non-randomized trials as part of the results.

Likewise, in the single-cohort studies, those patients with intensive follow-up were almost twice as likely to have a curative resection for recurrence than patients in the historic control group, thus strongly advocating the benefit of a more intense follow-up strategy, which also allows a higher chance of picking up recurrence in those patients at higher risk.

However, as already discussed above, others advocate less intense follow-up, stating that the numbers needed to treat are far too high and that there is no adequate curative treatment for most recurrences anyway, especially if TME was performed correctly in the first instance. There is also new discussion as to who should be carrying out the follow-up process. Advocates of intense follow-up recommend surgical teams, and this seems logical in the immediate and short-term postoperative period. But in the long term there are questions of which healthcare provider should be the key contact. Should there be more primary care involvement for example? Which clinicians would provide quicker diagnosis of recurrence but also have a more positive effect on the quality of life of the patients? Such questions are indeed contentious; however, currently each patient situation should be considered individually as there is no real hard evidence to answer the above questions.

Prognosis

Studies such as Eurocare-3 have clearly shown how survival from colorectal cancer in the UK differs from other European countries, with overall 5 year survival being 45.1% for men and 49.6% for women.

Blood tests: serum tumour markers

CEA has been used extensively both in the USA and in Europe for follow-up for many years now, with the prospect that it is able to detect recurrence possibly up to 6 months prior to symptoms. There is evidence that this relatively simple test leads to earlier diagnosis and more reoperative management. But does this mean more cure for recurrence? This is unlikely, and there is little evidence to suggest that second-look surgery changes survival rates on the whole. It has been demonstrated in the literature that at times there may be hundreds of thousands of patients undergoing serial CEA measurements, and this undoubtedly has cost implications. The same study showed less than excellent sensitivities and specificities for CEA levels and recurrence and led the investigators to conclude that cancer cures attributable to CEA monitoring are, at best, infrequent. It is still questioned whether this small gain justifies the substantial financial cost and physical and emotional stress that this intervention may cause for patients.

However, a Dutch meta-analysis published more recently suggests the CEA should in fact be used primarily for the early diagnosis of recurrence. But this study also suggests that the numbers needed to treat are high, and so still may not be cost effective. These contentious issues may be resolved when the results of the St Marks randomized trial on CEA levels during follow-up are published.

Other markers have been studied, e.g. CA 19-9 (carbohydrate antigen) and TPA (tissue polypeptide) and compared with CEA, but CEA seems to be the most sensitive and combinations of measured markers do not seem to change overall sensitivity and specificity.

Flexible endoscopy/colonoscopy

Endoscopy is often performed at 1, 3 and 5 years following resectional surgery and has two important roles:

- recognition of recurrence
- detection of metachronous tumours.

It must be remembered that endoscopic techniques are relatively insensitive for detecting recurrences, but this is because half of all recurrences do not originate from within the bowel lumen. Thus its main use is that of diagnosing new primary neoplastic lesions. Guidelines set forth by the American Society of Colon and Rectal Surgeons recommend surveillance colonoscopy 1 year postoperatively, then every 3 years if the colon is normal, which implies annual examinations until a normal examination is obtained.

The data from a study reported the findings of annual surveillance colonoscopy in colorectal cancer patients, and concluded the following.

- The risk for a positive initial postoperative colonoscopy was 18.3%.
- After a normal colonoscopy, risk for finding a neoplastic polyp on the next annual examination was 10.4%.
- After a positive colonoscopy, risk for finding a neoplastic polyp on the next annual examination was 43%.
- The risk for a positive examination after finding multiple neoplastic polyps was 70.4%.

The investigators concluded that annual follow-up colonoscopy for 2 years after colorectal cancer surgery is beneficial for detecting recurrent and metachronous neoplasms and that the interval between subsequent examinations may be increased depending on the result of the most recent examination.

Imaging

Ultrasonography has long been used to image the liver accurately with regards to metastases. Generally it is sensitive and specific for even small lesions (e.g. 1 cm in diameter), although it is highly operator dependent.

Besides imaging the liver, ultrasound scanning has also been used to assess local recurrence of rectal cancer using endoluminal ultrasound techniques. A study showed the sensitivity of endoluminal ultrasound and CT for recurrence to be 100% and 86% respectively with a specificity of 89% and 93% respectively. Overall, the accuracy for both imaging techniques was found to be 93% and 83% respectively. However, because of small study size the findings were not statistically significant.

CT has long been considered to be effective in the diagnosis of colorectal cancer recurrence but with varying sensitivities. It undoubtedly, along with MRI, is excellent at delineating recurrence in the surrounding pelvic cavity. In many centres it is the diagnostic tool of choice. But its limitations must be remembered:

- lesion diameter
- low sensitivity for lymph node recurrence
- artefacts from surgical clips
- difficulty in differentiating between postoperative scarring and recurrence.

MRI has proved to be superior to CT tissue characterization even in the presence of surgical artefact enabling just as high, if not higher, sensitivities and specificities compared with CT. However, availability, cost and patient factors (e.g. metal prosthesis, claustrophobia) make this investigation less usable on occasion. MRI may also be used endoluminally and this may yet yield better results, possibly because of its deeper focal length.

One of the difficulties of both CT and MRI is differentiating between recurrence and scar tissue; however, this issue may be improved by the use of serial scanning rather than just one-off imaging. PET may also be used in allowing differentiation of recurrence from surgical change.

Screening for colorectal cancer

It is well established that 'early' colorectal tumours (i.e. stage A) have a much better prognosis than more advanced tumours, and it therefore makes sense to try to diagnose colorectal cancer at an early stage. Unfortunately, the common symptoms of colorectal cancer (i.e. change of bowel habit, abdominal pain and rectal bleeding) only occur when the tumour is relatively locally advanced. The only way in which to reliably identify cancers at an early stage is to employ population screening, and a great deal of research effort has gone into this approach over the past 20 years.

There are now three population-based randomized controlled trials which have demonstrated that screening asymptomatic populations with faecal occult blood testing can reduce disease-specific mortality by 15–30%. Unfortunately, standard faecal occult blood testing, which is positive in about 2% of an unselected population between the ages of 50 and 75 years, is only about 50% sensitive. Thus a negative faecal occult blood test is not a guarantee that colorectal cancer is not present, although a positive test is associated with a 50% chance of colorectal neoplasia and a 12% chance of having invasive malignancy. It is possible to increase the sensitivity of faecal occult blood testing, but this reduces the specificity to such an extent that a large number of negative investigations would have to be carried out.

Another approach to screening is to use flexible endoscopy as the primary screening test. To use colonoscopy in this way would be impractical, but flexible sigmoidoscopy is a more realistic proposition. As 75% of colorectal cancers are within the reach of this instrument and as the presence of an adenomatous polyp in the left colon is an indicator of possible neoplasia on the right of the colon, then a 'once-only' flexible sigmoidoscopy may be a useful screening test. This is the subject of a randomized trial carried out in the UK, but mortality data from this trial are not yet available. One of the problems with all forms of colorectal cancer screening is the low compliance rate, which is presumably related to a mixture of ignorance, fear and distaste.

Large bowel polyps

Polyps

The term polyp is rather imprecise and in its broadest sense can be taken to mean a protuberant growth which can be either benign or malignant. As far as the colon and rectum are concerned, however, the term is usually taken to mean a benign swelling arising from the colonic or rectal mucosa, although, as it will be seen later on, certain types of polyp may contain a malignant or invasive focus and may indeed be an essential precursor in the development of colorectal cancer. Colorectal polyps may be inflammatory, hamartomatous, metaplastic or adenomatous. These will be dealt with in turn, but particular attention will be paid to the adenomatous variety in view of their close association with colorectal cancer.

Inflammatory polyps

These occur in ulcerative colitis, Crohn's colitis, diverticulitis, chronic dysentery and in benign lymphomatous lesions of the colon. They are sometimes referred to as pseudopolyps as they are commonly formed from an island of hypertrophied mucosa in an area of inflammation and ulceration. These polyps tend to be small, rarely exceeding 0.5 mm in diameter and consist of inflamed congested mucosa with oedematous changes in the submucosa.

Hamartomatous polyps

Hamartomatous polyps may be found in two forms, as juvenile polyps and as the familial Peutz–Jeghers syndrome. Juvenile polyps are found in infants or children and are often multiple, being round or oval with a smooth surface. At the time of diagnosis most lesions are pedunculated with a transition from normal colonic mucosa to a type of glandular tissue at the junction of the stalk and the polyp. The polypoidal substance consists of vascular tissue infiltrated by inflammatory cells and contains cystic spaces maintained by mucus-secreting columnar cells. There is a familial tendency in juvenile polyposis with the majority of patients presenting before the age of 10 years. Male children predominate over female. Fortunately, they are single in 70%, and 70% occur in the rectum and distal sigmoid colon.

The polyps occurring in the Peutz–Jeghers syndrome are associated with pigmented lesions (a bluish brown discoloration) on the face and on the lingual and buccal mucosa. Here the familial tendency is very strong. The polyps are almost always multiple and are found more commonly in the small bowel than in the colon or rectum. On histological examination the basic malformation is found in the muscularis mucosae. Unlike juvenile polyposis there is a significant malignant potential and there are reports of carcinoma arising in young patients with this syndrome (see also Chapter 29).

Metaplastic polyps

These are generally plaque-like growths which vary in size from 1 to 2 mm but rarely exceed 5 mm in maximum diameter. Although they are most commonly found in the rectum, the whole of the large bowel is susceptible. There is no specific age distribution or predisposing factor. On histological examination there is lengthening of the mucosal glands with dilatation of the goblet cells and evidence of inflammatory infiltration of the lamina propria. It is not understood why these lesions arise but they are very rarely symptomatic and do not appear to be premalignant. There is however some evidence that the presence of metaplastic polyps in the rectum or distal colon may be associated with an increased risk of adenomatous polyps or even carcinoma in the more proximal colon.

Adenomatous polyps

Pathology

Adenomatous polyps are benign neoplastic growths arising from the mucosa of the intestine and although they may occur anywhere between the stomach and the rectum, they are most common in the large intestine. In the Western world they are extremely common and postmortem studies indicate that they are found in more than 30% of people over the age of 60 years. The distribution of polyps is similar to that of adenocarcinoma, i.e. commonest in the rectum and left side of the colon, rare in the transverse colon and with a slight increase in the incidence in the right side of the colon and caecum. Adenomas are highly variable in size and macroscopically may be pedunculated or sessile. Recently the concept of the flat adenoma, which can be defined as an area of adenomatous change barely discernible macroscopically has emerged, but the significance of these lesions has yet to be established. In the colon, adenomas are normally pedunculated whereas in the rectum they are commonly sessile. The villous papilloma is a sessile adenoma made up of frond-like strands which grows as a carpet on the rectal mucosa (Figure 30.50).

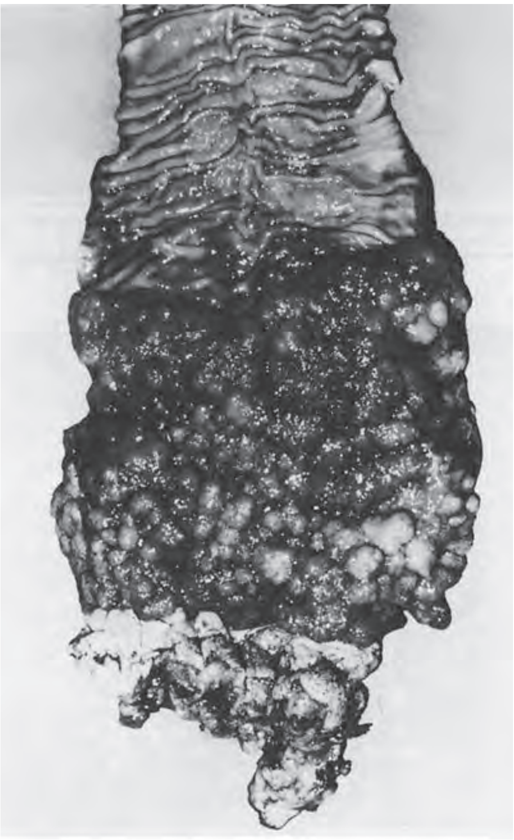


Figure 30.50 Extensive villous adenoma of the rectum.

Histologically, the epithelium in an adenoma can be arranged in tubular pattern consisting of closely packed glands or a villous pattern where the epithelial cells are arranged on frond-like extensions from the surface of the tumour. In practice the majority of adenomatous polyps display a mixture of tubular and villous patterns and can be described as tubulovillous. When the pathologist examines an adenoma it is important to establish whether or not there are any areas of invasion where dysplastic cells have transgressed the basement membrane into the fibrous stalk of the polyp. This will be found in about 50% of all adenomas that are over 2 cm in maximum diameter. In general, villous adenomas are more likely to undergo malignant change than tubular adenomas but this is by no means an absolute rule.

Aetiology

The aetiology of adenomatous polyps is essentially the same as that of colorectal cancer, indeed it is believed that the majority of colorectal cancers arise from pre-existing adenomatous polyps (see next section). In summary, although environmental factors (probably mainly dietary) have important implications for the formation of polyps, the genetic background is crucial. Not only are there dominantly inherited mutations which predispose to the development of polyps and cancer (see sections Familial adenomatous polyposis and Hereditary non-polyposis colorectal cancer) but there are also more subtle genetic variations which have an important impact on the predisposition to develop adenomatous polyps. This will be dealt with in detail in the section Colorectal cancer.

Clinical features

Most adenomatous polyps are asymptomatic and the diagnosis is made on routine examination. Nevertheless, both occult and frank bleeding can occur and patients may present with either rectal bleeding or anaemia. Occasionally, polyps may be extruded from the anal canal and may be misdiagnosed as prolapsing haemorrhoids. The retrograde propulsion of larger pedunculated polyps may produce abdominal pain and in extreme cases lead to the development of colocolic intussusception.

Rectal polyps may be accompanied by tenesmus and a change in bowel habit to diarrhoea. This may be the result of mucoid discharge from the surface of the polyps. This feature is particularly common with villous papillomas where spurious diarrhoea from the abundant mucus discharge leads to a failure in health, dehydration and electrolyte disturbance. In the mucus, sodium and chloride concentrations are similar to plasma but the potassium concentrations are between three and 20 times greater. Thus in larger papillomas, hypokalaemia and metabolic acidosis may result in lethargy, muscle weakness, mental confusion and in some extreme cases renal failure. These metabolic disturbances require attention prior to any attempt at surgical treatment.

Diagnosis

The diagnosis of adenomatous polyps is made on either large bowel endoscopy or barium enema. Large bowel endoscopy (colonoscopy in particular) is more sensitive at identifying polyps but large polyps can easily be seen on a high-quality, double-contrast barium enema (Figure 30.51). In some instances, it may be difficult to distinguish among a polyp, a diverticulum and faecal material, and this gives rise to a significant false-positive rate of polyp detection using barium enema. It must also be stressed, however, that colonoscopy is not 100% sensitive either and it has been demonstrated that repeat colonoscopy can often demonstrate lesions which were missed on the previous colonoscopy and miss lesions which had already been seen. One advantage of endoscopic diagnosis of polyps is that a biopsy can be taken of the larger polyps. This is important as it is often difficult to distinguish between a large benign adenomatous polyp and a polypoid carcinoma. Colonoscopy also offers the opportunity to remove the polyp using snare diathermy (Figure 30.52).

Prognosis

The prognosis of benign polyps in themselves is usually good, with the proviso that large villous adenomas can be extremely

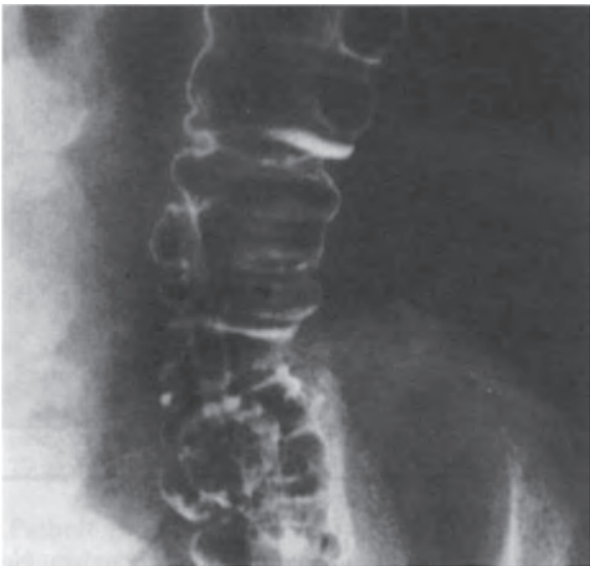


Figure 30.51 A 2.0 cm polyp on a stalk demonstrated by double-contrast barium enema.

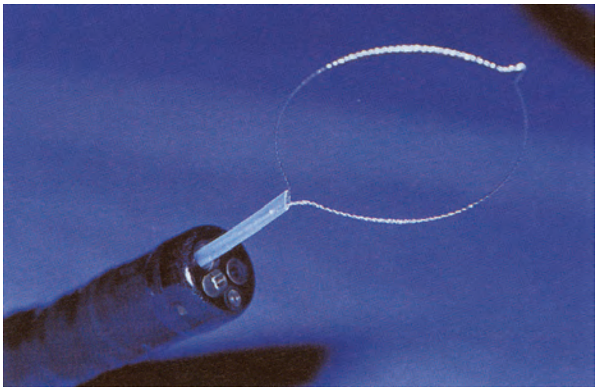


Figure 30.52 Colonoscopic snare.

debilitating (see above). However, as most carcinomas are thought to arise from polyps, the potential for malignant change should never be ignored and wherever it is feasible benign adenomatous polyps should be removed.

Treatment

The mainstay of treatment of adenomatous polyps is endoscopic polypectomy. Using a colonoscope, a wire loop is placed around the stalk of a pedunculated polyp; a blended current is then passed along the snare which is gradually tightened until it cuts through the stalk (Figure 30.53). The polyp should then be retrieved for histological examination. Occasionally, polypectomy may be accompanied by brisk arterial bleeding. If this occurs, the endoscopist should grasp the bleeding stalk with the snare, tighten the snare and hold in place for about 5 minutes without using diathermy current. This will usually stop the bleeding.

Another important complication of polypectomy is perforation of the colon. All patients undergoing colonoscopy with polypectomy should be warned of the risks of bleeding and perforation, as both may require emergency surgery. When a colonic polyp is sessile or has a very broad stalk, then it may be hazardous to carry out a standard polypectomy owing to the risk of perforation. In this case it is possible to elevate the polyp by injecting saline into the submucosa (Figure 30.54). This will allow much safer polypectomy. In very large polyps, it may be possible to carry out piecemeal snare excision. This is a difficult and hazardous procedure and should only be carried out by experienced endoscopists.

When a polyp has been completely excised and sent for histology, a focus of invasion will occasionally be found within the polyp making it a polyp cancer. It is generally agreed that if the polyp has been removed in one piece and the invasion does not extend to the resection margin then further surgery is not required. However, if the resection margin is involved, the patient should then be offered a colectomy. For this reason, if an endoscopist is concerned that a polyp may in fact be a cancer, it is useful to mark the site of excision using an intramural injection of India ink via the colonoscope. Having said this, many surgeons will take the view that a large polyp in the colon is highly likely to be malignant, even in the absence of confirmatory biopsies and will advise the patient to go straight to colectomy.

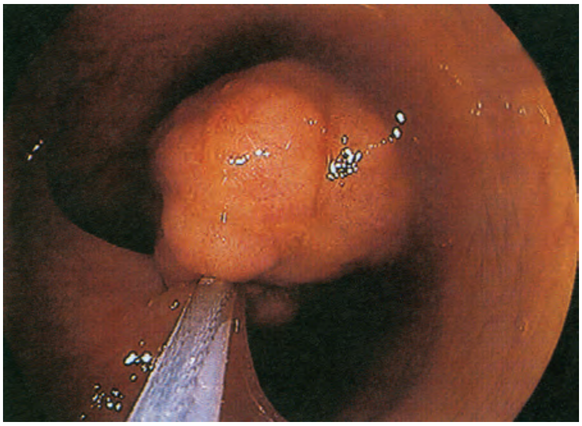


Figure 30.53 Snare excision of a polyp.

Although these endoscopic approaches are suitable for the majority of colonic polyps, rectal polyps pose a different problem. Occasionally, a large pedunculated prolapsing rectal polyp can be pulled out through the anal canal and the stalk simply ligated and divided. However, the majority of rectal polyps are sessile and require a different approach. Traditionally a low sessile rectal polyp was treated by submucosal saline injection to lift it away from the muscle wall and then transanal excision using Park's anal retractors to gain access (Figure 30.55).

For polyps which are higher in the rectum, however, this is often not feasible, and for this reason TEMS was developed. This employs a sophisticated operating sigmoidoscope with

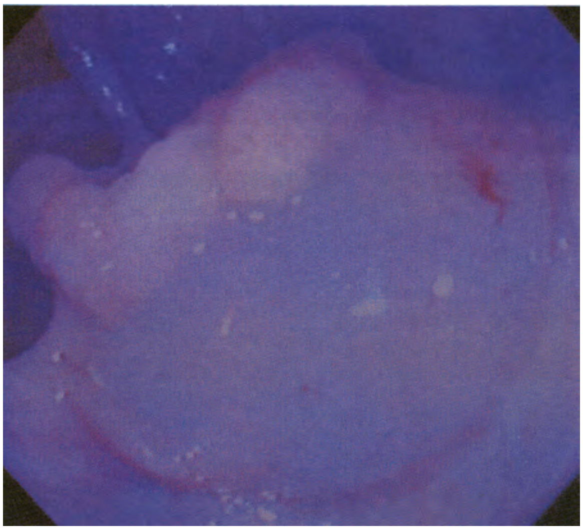


Figure 30.54 Elevation of polyp after submucosal saline injection.

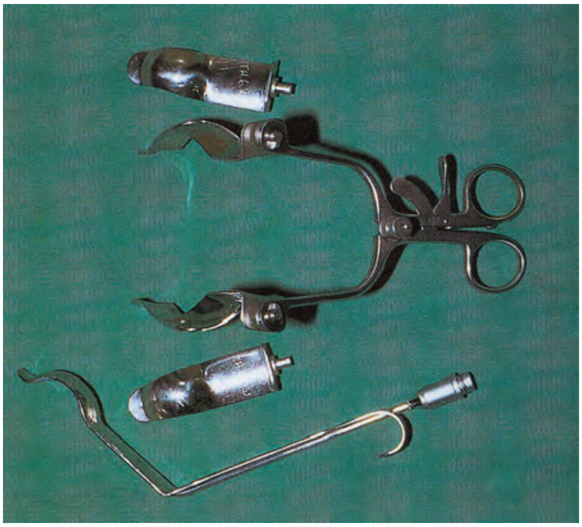


Figure 30.55 Parks' anal retractor for local transanal excision of rectal tumours.

a binocular optical system (Figure 30.48), which, by means of continuous insufflation and suction, offers an excellent operating environment to remove such polyps. When these polyps are situated posteriorly, it is safe to remove a full-thickness disc of rectal wall and this is a sensible precaution as a proportion of such polyps will harbour invasive malignancy. However, anteriorly above about 10 cm this is hazardous as it may lead to perforation into the peritoneal cavity and a submucosal technique must be employed. An alternative procedure is the trans-sphincteric approach described by Yorke Mason, which involves formal division of the anal sphincter mechanism in order to gain access to the rectum posteriorly. Owing to the currently available sophisticated techniques for transanal excision, however, this approach is seldom used nowadays. Occasionally, a patient will have such an extensive carpet of adenoma throughout the rectum that a transanal approach is not feasible. In this case a total proctectomy (i.e. an anterior resection with a mucosectomy down to the dentate line) has to be carried out. Continuity can then be restored by means of a sutured coloanal anastomosis.

Familial adenomatous polyposis

FAP is a dominantly inherited condition characterized by the appearance of multiple adenomatous polyps throughout the colon, which inevitably lead to colorectal cancer by the fourth decade of life. It is caused by a germline mutation of the *APC* gene, which is located on chromosome 5q and functions as a tumour suppressor gene with implications for cell-to-cell adhesion. The severity of the polyposis is variable and is related to the specific mutation, e.g. mutations at codon 1309 (exon 15) are associated with a very dense polyposis.

Although colonic polyps are the most important phenotypic manifestation of this condition, there are also extracolonic features which appear to a varying degree. These include adenomas and carcinomas of the small bowel and particularly the periampullary area of the duodenum, congenital hypertrophy of the retinal pigment epithelium (CHRPE) (Figure 30.56) and desmoid tumours. When FAP occurs in association with desmoid tumours, epidermoid cysts and bony exostoses, it is termed Gardener syndrome and this appears to occur in about 10% of cases.

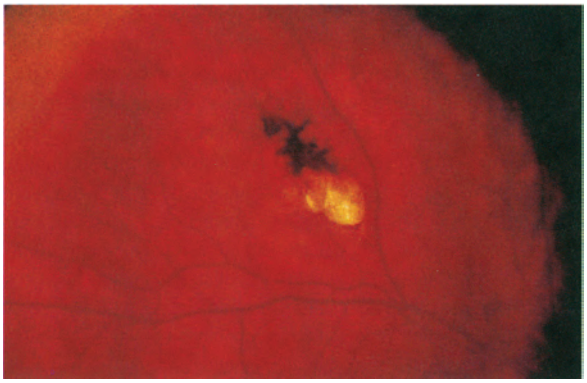


Figure 30.56 Congenital hypertrophy of the retinal pigment epithelium.

Clinical features

Commonly this condition is picked up by surveillance, as children of affected individuals have a one in two chance of developing colonic polyps. However, in an individual not suspected of having FAP, the most usual presenting symptom is rectal bleeding. CHRPEs may be seen on retinoscopy and desmoid tumours usually present as abdominal wall or intra-abdominal masses.

Diagnosis

The age at which the polyps develop is variable, but it is usually in the late teenage years and early twenties. For this reason, individuals at risk of FAP should undergo yearly sigmoidoscopy from puberty onwards. Genetic screening is now possible, and once the specific mutation in the *APC* gene has been detected by genetic sequencing in an affected individual, his or her children can then be screened for that specific mutation. If they do not carry the mutation, then endoscopic surveillance will not be necessary.

Treatment

The treatment for FAP is colectomy soon after the development of the adenomatous polyps (usually in the early twenties). There is some debate whether a total colectomy with ileostomy or ileoanal pouch should be performed or whether it is safe to carry out a subtotal colectomy with ileorectal anastomosis followed by careful follow-up of the rectum with ablation of the polyps when they arise. Mutation analysis may help in this respect and it might be reasonable to plan the operation depending on the likely density of polyps (i.e. those with the mutation at codon 1309 at exon 5 might be best served by a total proctocolectomy, whereas in other mutations an ileorectal anastomosis may be appropriate). More work is required in this area.

Hereditary non-polyposis colorectal cancer

While FAP probably accounts for 1–3% of all colorectal cancers, HNPCC may account for 5–10%. Like FAP, HNPCC is a dominantly inherited condition, but as it does not lead to multiple polyposis it is more difficult to recognize. The germline mutations which lead to HNPCC are thought to occur in DNA mismatch repair genes and currently four are recognized – *hMSH2*, *hMLHl*, *hPMSl* and *hPMS2* – but 90% of mutations in HNPCC are found in *hMSH2* and *hMLHl*.

Clinically, the diagnosis of HNPCC depends on the 'Amsterdam criteria', which require three or more relatives with histologically verified colorectal cancer, one of whom is a first-degree relative of the other two, with at least one patient being less than 50 years of age. HNPCC is associated with an excess risk of extracolonic cancers and is classified into 'Lynch' syndromes 1 and 2 on the basis of extracolonic disease. In Lynch syndrome 2, there is a particularly strong association with endometrial and ovarian cancer but other malignancies which may occur include transitional cell carcinoma of the ureter and renal pelvis, gastric cancer, pancreatic cancer, biliary tract cancer and haematological cancers.

Clinical features

Clinical features are merely those of the cancer that develops, particularly colorectal cancer. It should be stressed that multiple polyps do not occur, although small numbers of adenomatous polyps are a feature of this condition. The lifetime risk for colorectal cancer is about 85% of affected individuals, which indicates incomplete penetrance of this inherited condition.

Diagnosis

In individuals at risk the current recommendations are that colonoscopy should be performed at the age of 25 years, then repeated every second year until the age of 35 and then repeated yearly. The reason for the increased frequency at 35 years is that the average age of developing colorectal cancer is around 40. Women with Lynch 2 syndrome may also be offered screening of the endometrium by endometrial aspiration and yearly measurement of the CA 125 tumour marker. It is now possible to carry out mutation analysis for the known mismatch repair gene mutations. Once a specific mutation is known in an affected individual, their offspring may be offered genetic screening.

Treatment

When a cancer has been identified it is usual to offer a subtotal colectomy and ileorectal anastomosis in view of the high risk of metachronous tumours. Whether or not prophylactic colectomy should be carried out in subjects known to be carrying a germline mutation in one of the appropriate genes is controversial but has its supporters. The main argument against this approach is the fact that the condition is not fully penetrant.

Unusual colorectal cancers

Squamous cell carcinoma

Squamous cell carcinoma of the colon or rectum is extremely rare and may be related to squamous metaplasia of existing adenomas or carcinomas. Chronic ulcerative colitis, schistosomiasis and radiotherapy are all other possible predisposing factors. This type of cancer presents as ordinary adenocarcinoma and the diagnosis is only made on histology. Squamous carcinoma of the colon and rectum must be differentiated from anal carcinoma which is a distinct clinical entity.

Carcinoid tumours

The most common site for carcinoid tumours is the appendix followed by the small intestine. The large bowel is the third most common site, with the rectum or anus being the most frequent. The tumours are usually polypoid in nature and appear to be slow growing, remaining asymptomatic for long periods of time. Indeed about 50% of patients with colonic carcinoids have no symptoms attributable to the primary tumour at the time of diagnosis. Carcinoids are dealt with in more detail in Chapter 29.

Lymphoma

Malignant lymphoma is the third most common malignant tumour of the large bowel after adenocarcinoma and carcinoid tumours but probably represents only about 1.5% of all colonic tumours. The commonest sites are the rectum and the caecum. The tumours are usually of the non-Hodgkin's variety. Because these tumours are rare, definitive guidelines on treatment are difficult to provide. However, it is generally recommended that colonic tumours should be treated by chemotherapy alone, whereas rectal tumours should be treated by a combination of radiotherapy and chemotherapy. Surgical resection should follow unless there is a complete response to non-surgical therapy.

Sarcoma

Various types of sarcoma can arise in the colon and rectum and present in very much the same way as adenocarcinoma. Of particular note in recent years is Kaposi's sarcoma which is a feature of HIV infection. This tumour has been reported as being the presenting feature in AIDS in about 25% of cases and is manifested by multiple lesions in the gastrointestinal tract and the skin. The rectum is a common site and this leads to symptoms suggestive of IBD (diarrhoea and ulceration). Radiotherapy and chemotherapy may be used to treat Kaposi's sarcoma and occasionally surgery is required.

Vascular disorders of the colon

Vascular disorders of the colon can be classified into three main categories:

- ischaemic colitis
- angiodysplasia
- visceral artery aneurysms.

Ischaemic colitis

The term ischaemic colitis can be used to cover ischaemic disorders of the large bowel caused by a number of different factors (Box 30.1). Generally speaking, ischaemia of the colon is caused by insufficient blood flow through the mesenteric vessels due to either thrombosis or embolus. Thrombosis of arteries to the colon is usually due to progressive narrowing of the vessels by atherosclerosis and may be accompanied by the development of a significant collateral blood supply. On the other hand, embolic occlusion of colonic vessels is usually sudden and therefore does not allow the development of collateral flow. In some cases ischaemia may occur in the absence of occlusive arterial disease, e.g. in patients with shock or congestive cardiac failure and following aortic reconstruction with ligation of the inferior mesenteric artery. This last cause is very unusual, however, as under most circumstances the entire colon can be adequately supplied by the superior mesenteric artery by way of the marginal artery. Various forms of vasculitis may also be responsible.

BOX 30.1 Causes of colonic ischaemia

Thrombosis – arterial or venous
- Arteriosclerosis
- Polycythaemia vera
- Portal hypertension
- Malignant disease of the colon
- Hyperviscosity syndrome due to:
  - platelet abnormalities
  - high molecular weight dextran infusion

Embolic from
- Left atrium (atrial fibrillation)
- Left ventricle (myocardial infarction)
- Atheromatous plaque in aorta

Vasculitis
- Polyarteritis nodosa
- Lupus erythematosus
- Giant cell arteritis (Takayasu's arteritis)
- Buerger disease
- Henoch–Schönlein disease

Surgical trauma to vessels
- Aortic reconstruction
- Resection of adjacent intestine

Non-occlusive ischaemia
- Shock – hypovolaemic or septic
- Congestive heart failure

'Spontaneous' ischaemic colitis

The degree of ischaemia in ischaemic colitis is highly variable and there are three clinical patterns:

- gangrenous ischaemic colitis
- transient ischaemic colitis
- stricturing ischaemic colitis.

Clinical features

Gangrenous ischaemic colitis

These patients are almost always elderly and usually have concurrent disease such as sepsis or cardiac failure. There is usually a history of acute onset of intense abdominal pain usually on the left side accompanied by bloody diarrhoea. On examination, the patient is severely ill and the abdomen is usually extremely tender with peritonism.

Transient ischaemic colitis

In this less dramatic form of the disease, patients are usually middle aged and may have evidence of peripheral vascular disease. There is usually a history of abdominal pain of several days' duration, and rectal bleeding is almost invariable. On examination, there is mild to moderate abdominal tenderness.

Stricturing ischaemic colitis

In these patients the clinical presentation may be insidious, although there may be a history of abdominal pain and rectal bleeding. Typically the patient will develop chronic abdominal discomfort and change of bowel habit and on examination there may be very little to find.

Diagnosis

Gangrenous ischaemic colitis

Plain abdominal radiographs in patients with colonic infarction show a blending of the mucosal folds, air in the wall of the colon and occasionally air in the portal venous system. If the colon has gone on to perforation, air under the diaphragm may be seen on an erect chest radiograph. Further investigation is usually not indicated as the patient will require urgent surgical intervention.

Transient ischaemic colitis

Here the abdominal radiograph will show 'thumb printing', which consists of multiple impressions of air due to submucosal oedema and haemorrhage. This is more clearly seen on a double-contrast barium enema. If colonoscopy is carried out, haemorrhagic oedema of the colonic mucosa will be seen. However, owing to the possibility of perforation, colonoscopy is not to be recommended if the diagnosis of ischaemic colitis is suspected.

Stricturing ischaemic colitis

Here the barium enema may reveal a relatively smooth stricture of the colon and on endoscopy thickened mucosa will be seen. Biopsy will show damage to the underlying muscularis propria associated with fibrosis.

Treatment

Gangrenous ischaemic colitis

When gangrenous ischaemic colitis is diagnosed then urgent laparotomy after adequate resuscitation is mandatory. The infarcted segment should be resected and the healthy bowel ends brought out as a colostomy or ileostomy and mucus fistula. Because of the gross contamination which often accompanies this condition the mortality is high.

Transient ischaemic colitis

If laparotomy is not indicated, treatment consists of general supportive measures, including intravenous fluids, parenteral nutrition where indicated and blood transfusion if the haemoglobin drops.

Stricturing ischaemic colitis

The treatment of this condition depends on symptoms and a colonic resection with primary anastomosis may be indicated.

Volvulus

Colonic volvulus or twisting of the colon is a relatively uncommon cause of large bowel obstruction in Western society, although it is much more usual in developing countries. The sigmoid colon is by far the most common site followed by the caecum. Very occasionally, volvulus of the transverse colon may occur.

Sigmoid volvulus

Sigmoid volvulus tends to affect elderly males, and typically the patient will be institutionalized with a previous history of chronic constipation, laxative dependency and unrelated medical problems.

Pathophysiology

An anticlockwise torsion of 180° is considered to be physiological. It may cause no symptoms and usually reverts spontaneously. However, if the posterior loop gradually fills with gas and stool and becomes heavier, it then changes its position and falls anterior to the empty loop, leading to a 360° anticlockwise torsion. Sigmoid volvulus may affect an individual for many years in a subacute recurring form but acute sigmoid volvulus is a surgical emergency because of the tight compression of the mesocolic vessels and massive distension of the colonic lumen.

Aetiology

There are probably many aetiological factors but the most important is probably a long sigmoid loop with a narrow mesentery leading to a predisposition to torsion. In addition, chronic constipation, a high-fibre diet and systemic neurological disease are important. In South America there is an association with Chagas disease.

Clinical features

In its chronic form the patient gives a history of intermittent lower colicky abdominal pain with abdominal distension which is relieved by the passage of flatus and loose stool. In the acute form, the patient will present as a large bowel obstruction with colicky abdominal pain, constipation and often massive distension. Vomiting is a late feature. On examination, the abdomen is distended and tympanic and may be tender over the dilated loop of colon. Tinkling bowel sounds can be heard initially but they vanish with either secondary ileus or perforation. Digital rectal examination reveals an empty distended rectum and there may be blood on the glove if gangrenous changes have set in.

Diagnosis

A plain abdominal radiograph (Figure 30.57) will confirm the diagnosis in 80% of cases. The diagnostic features include an enormously distended sigmoid colon arising from the left iliac fossa towards the right hypochondrium. It may lead to an elevated diaphragm. The loop has the appearance of a coffee bean and there is commonly a bird beak deformity at the site of the torsion. In late cases, peritoneal fluid is indicated by a ground-glass appearance, and there may be progressive distension of the small intestine with gas and fluid.

Complications

Acute sigmoid volvulus is complicated by gangrene and perforation of the twisted segment because of the tight compression of the mesocolic vessels and the massive distension. This may lead to faecal peritonitis. Occasionally, the massive distension may also lead to respiratory embarrassment.

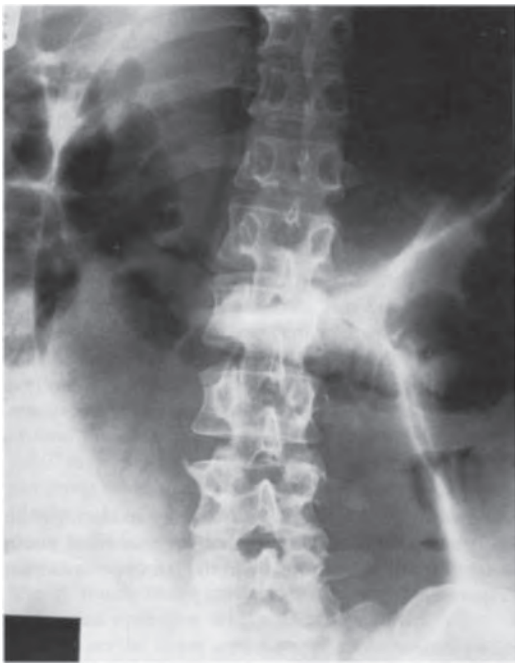


Figure 30.57 Plain abdominal radiograph of acute sigmoid volvulus.

Treatment

The first step should be to correct any fluid and electrolyte imbalance with intravenous replacement and then to attempt endoscopic reduction. This may be achieved by either a rigid sigmoidoscope or a flexible colonoscope and is successful in 80% of cases. A flexible instrument is preferable as the point of torsion may often be higher than the limit of the sigmoidoscope. Colonoscopy also allows for examination of the mucosal aspect for evidence of gangrene. If a flexible endoscope is not available and the rigid sigmoidoscope will not reach the point of torsion then it is possible to pass a blunt-ended flatus tube through the rigid sigmoidoscope to attempt a blind reduction.

It should be stressed that non-operative reduction does not constitute adequate treatment as the recurrence rate approaches 90%. Ideally, after successful endoscopic reduction, elective surgery should be carried out as quickly as possible as a one-stage sigmoid colectomy. However, if surgery is required urgently because of suspected strangulation or if non-operative decompression is unsuccessful, the standard treatment is resection of the sigmoid colon with a left iliac fossa colostomy and mucus fistula. Alternatively, the rectal stump may be oversewn as a Hartmann's procedure. In the relatively fit patient, however, it may be possible to do a sigmoid resection with on-table colonic irrigation followed by primary anastomosis but this should only be attempted when there is no doubt about the viability of the involved colon.

Caecal volvulus

The term caecal volvulus is really a misnomer as 90% of the cases are more accurately described as ileocolic torsion with only 10% involving the caecum alone. The propensity for caecal volvulus depends on a congenital abnormality of intestinal rotation in which the right colon does not fuse with the posterior abdominal wall but rather remains on a mesentery and is therefore highly mobile. It should be stressed that caecal volvulus is uncommon accounting for less than 1% of all cases of large bowel obstruction and only about 10% of all cases of volvulus.

Clinical features

There are three well-recognized modes of presentation. First, the patient may present with acute intestinal obstruction and this accounts for about 50% of cases. Second, the patient may present with rapid onset of acute abdominal pain and peritonitis secondary to gangrene of the affected segment of intestine. Third, a small number of patients present with chronic intermittent obstructive symptoms.

Diagnosis

Diagnosis in the acute stage is usually made on plain abdominal radiograph, which reveals a large air-filled viscus with an air – fluid level usually in the left upper quadrant (Figure 30.58).

Complications

The main complication as with sigmoid volvulus is gangrene leading to perforation and faecal peritonitis.

Treatment

Colonoscopic decompression is not usually successful and the mainstay of treatment is laparotomy with right hemicolectomy. Primary anastomosis is appropriate if there is no peritoneal soiling but if there is any doubt about the safety of an anastomosis, an ileostomy and mucus fistula should be fashioned. Another surgical option when the affected segment of bowel is completely viable is to carry out a caecopexy. This involves mobilizing the peritoneum of the right paracolic gutter to form a pouch into which the caecum can be inserted and sutured into place.

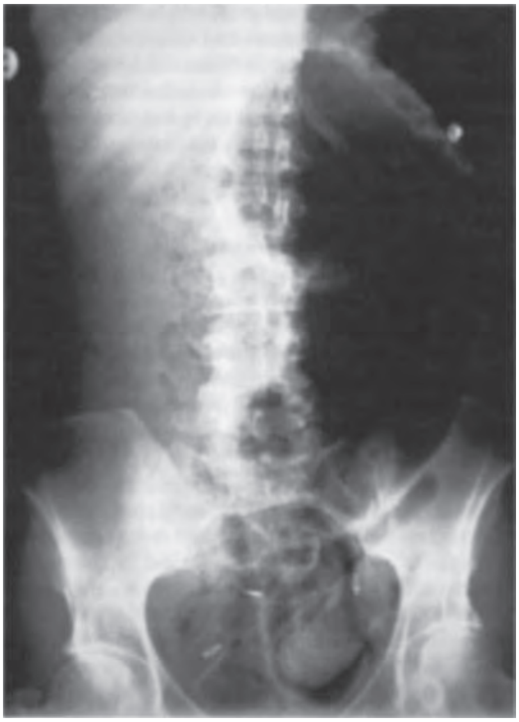


Figure 30.58 Plain abdominal radiograph of acute caecal volvulus.

Angiodysplasia

Angiomatous lesions of the bowel were first identified about 40 years ago with the introduction of angiography, and are now recognized to be relatively common. They characteristically occur in patients over the age of 60 years and are not usually associated with other angiomatous lesions of other viscera or of the skin. These lesions have been called angiodysplasias, angiomas, haemangiomas, arteriovenous malformations or vascular ectasias. Although they may occur anywhere in the gastrointestinal tract, they are commonest on the right side of the colon and histologically they consist of thin-walled vascular channels in the submucosa.

Aetiology

The cause of these lesions is unclear. They do not appear to be congenital and should be distinguished from hereditary haemorrhagic telangiectasia (Rendu–Osler–Weber syndrome). There are various theories as to the aetiology, including obstruction of submucosal veins as they pass through the muscle layers of the colonic wall and repeated episodes of bowel ischaemia with arteriovenous shunting. Both of these theories explain why the lesions should predominantly occur on the right side of the colon. This follows Laplace's law (tension is proportional to diameter and pressure) as the right colon has the widest diameter of the large bowel.

Clinical features

The clinical features of angiodysplasia depend on its tendency to bleed. Patients may present with anaemia or on some occasions with massive rectal bleeding.

Diagnosis

The main diagnostic modality is mesenteric angiography which characteristically shows an early draining vein parallel to a feeding artery (Figure 30.59). Colonoscopy may also detect the typical 'cherry red' spots which are characteristic of this condition (Figure 30.60).

Treatment

The most effective treatment for this condition is resection of the affected portion of bowel (usually a right hemicolectomy). In extreme conditions where the patient is actively haemorrhaging and the appropriate expertise is available, selective vasopressin infusion or transcatheter embolization with Gelfoam at arteriography may stop major haemorrhage. If the lesion is detected at colonoscopy then grasping it with diathermy forceps, lifting it away from the muscular wall and applying diathermy current may coagulate the abnormal vessels and prevent recurrent bleeding.

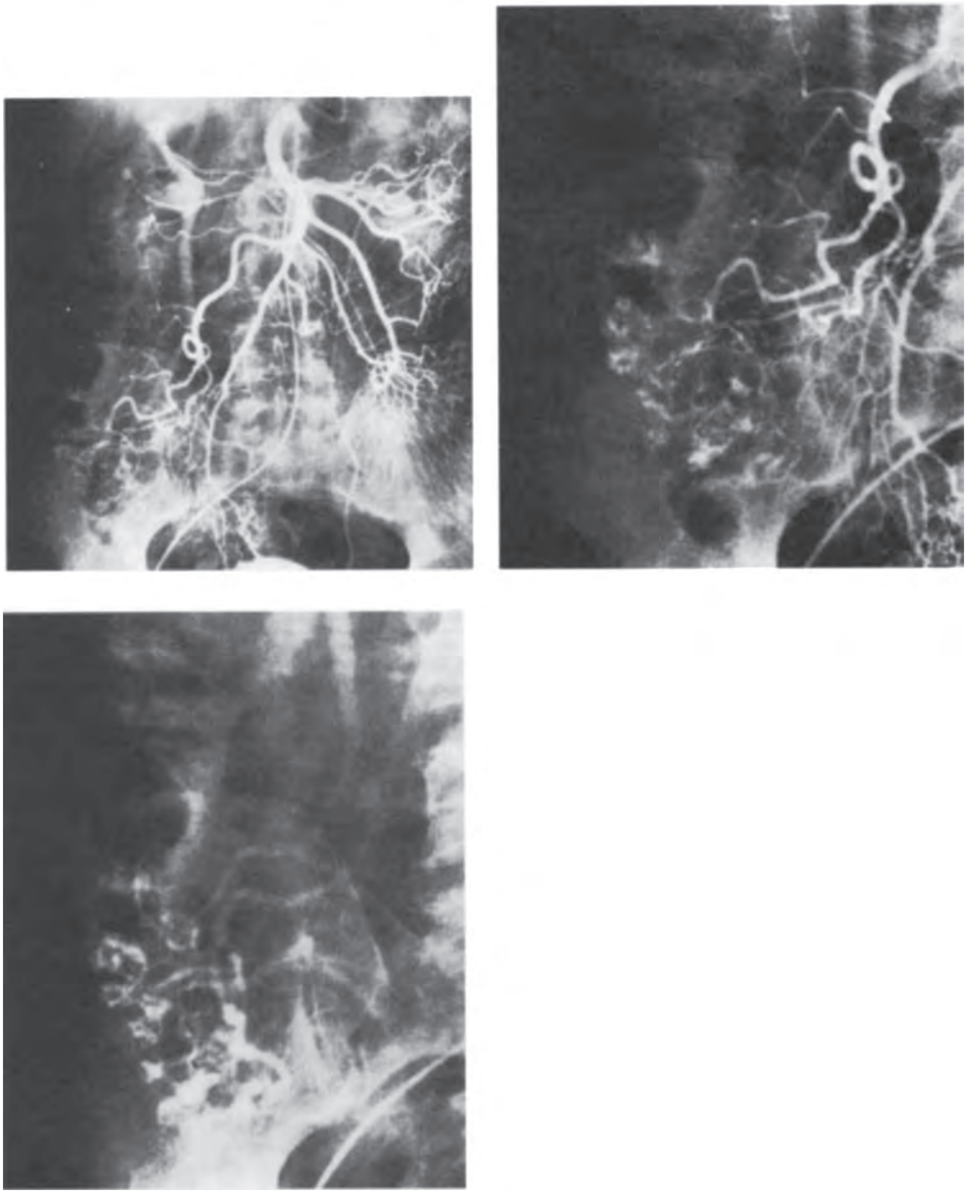


Figure 30.59 Angiographic appearances of angiodysplasia.

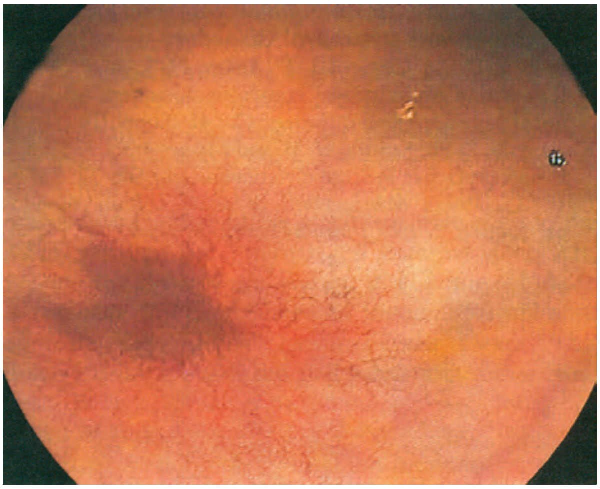


Figure 30.60 Colonoscopic appearances of angiodysplasia.

Visceral artery aneurysms

Aneurysms of the visceral arteries are relatively common but aneurysms of the colonic arteries are uncommon. There are various aetiological factors including congenital abnormalities, atherosclerosis, trauma, medial degeneration and forms of arteritis. Rupture of these aneurysms gives rise to severe abdominal pain and signs of internal haemorrhage. Successful treatment depends on prompt operative intervention.

GUIDE TO FURTHER READING

Functional disorders

Jayne DG, Schwandner O, Stuto A. Stapled transanal rectal resection for obstructed defecation syndrome: one-year results of the European STARR Registry. *Dis Colon Rectum* 2009;**52**:1205–1212; discussion 1212–1204.

Tan E, Anstee A, Koh DM, *et al.* Diagnostic precision of endoanal MRI in the detection of anal sphincter pathology: a meta-analysis. *Int J Colorectal Dis* 2008;**23**:641–651.

Tan E, Ngo NT, Darzi A, *et al.* Meta-analysis: sacral nerve stimulation versus conservative therapy in the treatment of faecal incontinence. *Int J Colorectal Dis* 2011;**26**:275–294.

Whitehead WE, Borrud L, Goode PS, *et al.* Fecal incontinence in US adults: epidemiology and risk factors. *Gastroenterology* 2009;**137**: 512–517.

Diverticular disease

Constantinides VA, Tekkis PP, Senapati A; Association of Coloproctology of Great Britain Ireland. Prospective multicentre evaluation of adverse outcomes following treatment for complicated diverticular disease. *Br J Surg* 2006;**93**:1503–1513.

Constantinides VA, Heriot A, Remzi F, *et al.* Operative strategies for diverticular peritonitis: a decision analysis between primary resection and anastomosis versus Hartmann's procedures. *Ann Surg* 2007;**245**:94–103.

Fearnhead NS, Mortensen NJ. Clinical features and differential diagnosis of diverticular disease. *Best Pract Res Clin Gastroenterol* 2002;**16**:577–593.

Kohler L, Sauerland S, Neugebauer E. Diagnosis and treatment of diverticular disease: results of a consensus development conference. The Scientific Committee of the European Association for Endoscopic Surgery. *Surg Endosc* 1999;**13**:430–6.

Tursi A, Papagrigoriadis S. The current and evolving treatment of colonic diverticular disease. *Aliment Pharmacol Ther* 2009;**30**:532–546.

Inflammatory bowel disease

Borody TJ, Campbell J. Fecal microbiota transplantation: current status and future directions. *Expert Rev Gastroenterol Hepatol* 2011;**5**:653–655.

Buffet-Bataillon S, Tattevin P, Sénéchal H, *et al.* Clostridium difficile 027 colitis: Hospital-onset but community-acquired. *Eur J Clin Microbiol Infect Dis* 2012.

Cornish JA, Teare J, Teoh TG, *et al.* A systematic review. the effect of restorative proctocolectomy on sexual function, urinary function, fertility, pregnancy and delivery. *Dis Col Rect* 2007;**50**:1128–1138.

Das P JM, Tekkis PP, Nicholls RJ. Risk of dysplasia and adenocarcinoma following restorative proctocolectomy for ulcerative colitis. *Colorectal Dis* 2007;**9**:15–27.

Delaney C, Fazio VW, Remzi FH, *et al.* Prospective, age-related analysis of surgical results, functional outcome, and quality of life after ileal pouch-anal anastomosis. *Ann Surg* 2003;**238**:221–228.

Jakobovits SL, Travis S. Infliximab for the treatment of ulcerative colitis: outcomes in Oxford from 2000 to 2006. *Aliment Pharmacol Ther* 2007;**25**:1055–1060.

Kelly CR, De Leon L, Jasutkar N. Fecal microbiota transplantation for relapsing Clostridium difficile infection in 26 patients: methodology and results. *J Clin Gastroenterol* 2012;**46**:145–149.

Lovegrove RE, Constantinides VA, Heriot AG, *et al.* A comparison of hand-sewn versus stapled ileal pouch anal anastomosis (IPAA) following proctocolectomy: a meta-analysis of 4183 patients. *Ann Surg* 2006;**244**:18–26.

Lovegrove R, Heriot AG, Constantinides V, *et al.* Meta-analysis of short-term and long-term outcomes of J, W and S ileal reservoirs for restorative proctocolectomy. *Colorectal Dis* 2007;**9**:310–320.

McLaughlin SD, Clark SK, *et al.* Tekkis PP, Review article: Restorative proctocolectomy, indications, management of complications and follow-up: a guide for gastroenterologists. *Aliment Pharmacol Ther* 2008;**27**:895–909.

Nelson R. Antibiotic treatment for Clostridium difficile-associated diarrhea in adults. *Cochrane Database Syst Rev* 2011;**9**:CD004610.

Olivas AD, Umanskiy K, Zuckerbraun B, Alverdy JC. Avoiding colectomy during surgical management of fulminant Clostridium difficile colitis. *Surg Infect (Larchmt)* 2010;**11**:299–305.

Penna C, Dozois R, Tremaine W, *et al.* Pouchitis after ileal pouch-anal anastomosis for ulcerative colitis occurs with increased frequency in patients with associated primary sclerosing cholangitis. *Gut* 1996;**38**:234–239.

Roberts S, Williams JG, Yeates D, Goldacre MJ. Mortality in patients with and without colectomy admitted to hospital for ulcerative colitis and Crohn disease: record linkage studies. *BMJ* 2007;**335**:1033–1041.

Tekkis PP, Heriot AG, Smith O, *et al.* Long-term outcomes of restorative proctocolectomy for Crohn's disease and indeterminate colitis. *Colorectal Dis* 2005;**7**:218–223.

Tekkis P, Tilney H, Sagar P, *et al.* Primary and salvage ileal pouch surgery in the United Kingdom – a multicenter study of 2,491 patients. *Dis Col Rect* 2007;**50**:756.

Thomas T, Abrams KA, Robinson RJ, Mayberry JF. Meta-analysis: cancer risk of low-grade dysplasia in chronic ulcerative colitis. *Aliment Pharmacol Ther* 2007;**15**:657–668.

Weston-Petrides G, Lovegrove RE, Tilney HS, *et al.* Comparison of outcomes after restorative proctocolectomy with or without defunctioning ileostomy. *Arch Surg* 2008;**143**:406–412.

Neoplasia

Al-Sukhni E, Milot L, Fruitman M, *et al.* Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis. *Ann Surg Oncol* 2012;**19**:2212–2223.

Aretz S, Uhlhaas S, Goergens H, *et al.* MUTYH-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. *Int J Cancer* 2006;**119**:807–814.

Basbug M, Arikanoglu Z, Bulbuller N, *et al.* Prognostic value of preoperative CEA and CA 19-9 levels in patients with colorectal cancer. *Hepatogastroenterology* 2011;**58**:400–405.

Bisgaard ML, Fenger K, Bulow S, *et al.* Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. *Hum Mutat* 1994;**3**:121–125.

Butterworth AS, Higgins JP, Pharoah P. Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. *Eur J Cancer* 2006;**42**:216–227.

Chan AT, Giovannucci EL, Meyerhardt JA, *et al.* Long-term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer. *JAMA* 2005;**294**:914–923.

Coleman MP, Gatta G, Verdecchia A, *et al.* EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. *Ann Oncol* 2003;**14**(Suppl. 5):v128–149.

Fearnhead NS, Wilding JL, Bodmer WF. Genetics of colorectal cancer: hereditary aspects and overview of colorectal tumorigenesis. *Br Med Bull* 2002;**64**:27–43.

Fedirko V, Tramacere I, Bagnardi V, *et al.* Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol* 2011;**22**:1958–1972.

Figueiredo JC, Levine AJ, Grau MV, *et al.* Colorectal adenomas in a randomized folate trial: the role of baseline dietary and circulating folate levels. *Cancer Epidemiol Biomarkers Prev* 2008;**17**:2625–2631.

Georgiou P, Brown G, Constantinides V, *et al.* Diagnostic accuracy of MRI in assessing tumor invasion within pelvic compartments in recurrent and locally advanced rectal cancer. *Dis Colon Rectum* 2009;**52**:9.

Georgiou P, Tan E, Gouvas N, *et al.* Extended lymphadenectomy versus conventional surgery for rectal cancer: a meta-analysis. *Lancet Oncol* 2009;**10**:1053–1062.

Grossmann I, Avenarius JK, Mastboom WJ, Klaase JM. Preoperative staging with chest CT in patients with colorectal carcinoma: not as a routine procedure. *Ann Surg Oncol* 2010;**17**:2045–2050.

Hardcastle JD, Chamberlain JO, Robinson MHE, *et al.* Randomized controlled trial of faecal occult blood screening for colorectal cancer. *Lancet* 1996;**348**:1472–1477.

Heald RJ. The 'Holy Plane' of rectal surgery. *J R Soc Med* 1988;**81**:503–508.

Hohenberger W, Weber K, Matzel K, *et al.* Standardized surgery for colonic cancer: complete mesocolic excision and central ligation: technical notes and outcome. *Colorectal Dis* 2009;**11**:354–364; discussion 364–365.

Jess T, Gamborg M, Matzen P, *et al.* Increased risk of intestinal cancer in Crohn's disease: a meta-analysis of population-based cohort studies. *Am J Gastroenterol* 2005;**100**:2724–2729.

John SK, George S, Primrose JN, Fozard JB. Symptoms and signs in patients with colorectal cancer. *Colorectal Dis* 2011;**13**:17–25.

Joshi AD, Corral R, Siegmund KD, *et al.* Red meat and poultry intake, polymorphisms in the nucleotide excision repair and mismatch repair pathways and colorectal cancer risk. *Carcinogenesis* 2009;**30**:472–479.

Kronborg O, Fenger C, Olsen J, *et al.* Randomized study of screening for colorectal cancer with faecal occult blood test. *Lancet* 1996;**348**:1467–1471.

Mackay SG, Pager CK, Joseph D, *et al.* Assessment of the accuracy of transrectal ultrasonography in anorectal neoplasia. *Br J Surg* 2003;**90**:346–350.

Mercury Study Group. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. *BMJ* 2006; **333**:779.

Mizoue T, Kimura Y, Toyomura K, *et al.* Calcium, dairy foods, vitamin D, and colorectal cancer risk: the Fukuoka Colorectal Cancer Study. *Cancer Epidemiol Biomarkers Prev* 2008;**17**:2800–2807.

Neri E, Faggioni L, Cerri F, *et al.* CT colonography versus double-contrast barium enema for screening of colorectal cancer: comparison of radiation burden. *Abdom Imaging* 2010;**35**:596–601.

Norat T, Bingham S, Ferrari P, *et al.* Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *J Natl Cancer Inst* 2005;**97**:906–916.

Pierre F, Tache S, Petit CR, *et al.* Meat and cancer: haemoglobin and haemin in a low-calcium diet promote colorectal carcinogenesis at the aberrant crypt stage in rats. *Carcinogenesis* 2003;**24**:1683–1690.

Rockey DC, Paulson E, Niedzwiecki D, *et al.* Analysis of air contrast barium enema, computed tomographic colonography, and colonoscopy: prospective comparison. *Lancet* 2005;**365**:305–311.

Rosman AS, Korsten MA. Meta-analysis comparing CT colonography, air contrast barium enema, and colonoscopy. *Am J Med* 2007;**120**:203–210.

Sandhu MS, White IR, McPherson K. Systematic review of the prospective cohort studies on meat consumption and colorectal cancer risk: a meta-analytical approach. *Cancer Epidemiol Biomarkers Prev* 2001;**10**:439–446.

Scott N, Jackson P, Al-Jaberi T, *et al.* Total mesorectal excision and local recurrence: a study of tumour spread in the mesorectum distal to rectal cancer. *Br J Surg* 1995;**82**:1031–1033.

SRCT (Swedish Rectal Cancer Trial). Improved survival with pre-operative radiotherapy in resectable rectal cancer. *N Engl J Med* 1997;**336**:980–987.

Sun Z, Wang PP, Roebothan B, *et al.* Calcium and vitamin D and risk of colorectal cancer: results from a large population-based case-control study in Newfoundland and Labrador and Ontario. *Can J Public Health* 2011;**102**:382–389.

Thun MJ, Namboodiri MM, Heath CW, Jr. Aspirin use and reduced risk of fatal colon cancer. *N Engl J Med* 1991;**325**:1593–1596.

Tsoi KK, Pau CY, Wu WK, *et al.* Cigarette smoking and the risk of colorectal cancer: a meta-analysis of prospective cohort studies. *Clin Gastroenterol Hepatol* 2009;**7**:682–688.

van Gijn W, Marijnen CA, Nagtegaal ID, *et al.* Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. *Lancet Oncol* 2011;**12**:575–582.

von Roon AC, Reese G, Teare J, *et al.* The risk of cancer in patients with Crohn's disease. *Dis Colon Rectum* 2007;**50**:839–855.

Weingarten MA, Zalmanovici A, Yaphe J. Dietary calcium supplementation for preventing colorectal cancer and adenomatous polyps. *Cochrane Database Syst Rev* 2008;**1**:CD003548.

Yeh CC, Sung FC, Tang R, *et al.* Polymorphisms of cytochrome P450 1A2 and N-acetyltransferase genes, meat consumption, and risk of colorectal cancer. *Dis Colon Rectum* 2009;**52**:104–111.

MAS colorectal cancer surgery

Clinical Outcomes of Surgical Therapy Study Group**.** A comparison of laparoscopically assisted and open colectomy for colon cancer. *N Eng J Med* 2004;**350**:2050–2059.

Color Study Group. COLOR: a randomized clinical trial comparing laparoscopic and open resection for colon cancer. *Dig Surg* 2000;**17**:617–622.

Jayne DG, Thorpe HC, Copeland J, *et al.* Five-year follow-up of the Medical Research Council CLASICC trial of laparoscopically assisted versus open surgery for colorectal cancer**.** *Br J Surg* 2010;**97**:1638.

Guillou PJ, Quirke P, Thorpe H, *et al.* Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): multicentre, randomised controlled trial. *Lancet* 2005;**365**:1718–1726

Kuhry E, Schwenk W, Gaupset R, *et al.* Long-term results of laparoscopic colorectal cancer resection. *Cochrane Database Syst Rev* 2008;**2**:CD003432.

Lacy AM, Garcia-Valdecasas JC, Delgado S, *et al.* Laparoscopy assisted colectomy versus open colectomy for treatment of non-metastatic colon cancer: a randomised trial. *Lancet* 2002;**359**:2224–2229.

Leung KL, Kwok SP, Lam SC, *et al.* Laparoscopic resection of rectosigmoid carcinoma: prospective randomized trial. *Lancet* 2004;**363:**1187–1192.

Liang J, Huang K. Oncologic results of laparoscopic versus conventional open surgery for stage II or III left-sided colon cancers: a randomized controlled trial. *Ann Surg Oncol* 2007;**14**:109–117.

Veldkamp R, Kuhry E, Hop WC, et al., COlon cancer Laparoscopic or Open Resection Study Group (COLOR). Laparoscopic surgery versus open surgery for colon cancer: short-term outcomes of a randomized trial**.** *Lancet Oncol* 2005;**6**:477–484.

Weeks JC, Nelson H, Gelber S, *et al*, Clinical Outcomes of Surgical Therapy (COST) Study Group. Short-term quality-of-life outcomes following laparoscopic-assisted colectomy vs open colectomy for colon cancer: a randomized trial. *JAMA* 2002;**287**:321–328.

Capsule colonoscopy

Fireman Z, Kopelman Y. The colon – the latest terrain for capsule endoscopy. *Dig Liver Dis* 2007;**39**:895–899.

Reliakim R, Fireman Z, Gralnek IM, *et al.* Evaluation of the PillCam Colon capsule in the detection of colonic pathology: results of the first multicenter, prospective, comparative study. *Endoscopy* 2006;**38**:963–970.

CHAPTER 31

Disorders of the anal canal

KEN CAMPBELL

Anatomy

The anal canal is about 3–4 cm long. It passes slightly posteriorly, starting at the anorectal angle and ending at the anal verge. In the male, the anal canal is related to the bulb of the urethra anteriorly, and in the female to the perineal body and the vagina anteriorly. Laterally, it is related to the ischiorectal fossa containing the inferior haemorrhoidal vessels and pudendal nerve (Figure 31.1); posteriorly lie the coccyx and the puborectalis muscle. As described below, the anal canal is surrounded by the internal and external sphincter muscles.

The epithelium of the anal canal is columnar above the anal valves and squamous below them; the site of the anal valves is also known as the dentate or pectinate line and this indicates the mucocutaneous junction. The anal valves consist of pits that represent the openings of the anal glands, and the glands themselves lie in the plane between the internal and external sphincters, helping to lubricate the anal canal. The mucosa above the dentate line is arranged in longitudinal columns covering the internal haemorrhoidal plexus. This mucosa is loose, but at the anal valves it becomes fixed to the internal sphincter by the mucosal suspensory ligament of Parks.

Strictly speaking the epithelium immediately above the anal valves is not columnar but cuboidal and this zone, which is between 0.5 and 2 cm in length, is known as the anal transition zone. Here there is a high density of sensory nerve endings. Although the epithelium is squamous below the dentate line there are no hair follicles, sebaceous glands or sweat glands in the anal canal. These appear only at the anal verge.

The internal anal sphincter is a downward, thickened extension of the circular muscle fibres of the rectum and extends approximately 1 cm below the anal canal. It consists entirely of smooth muscle and is innervated by the pelvic autonomic plexus. The external sphincter, on the other hand, is made up of skeletal muscle that is arranged around the anal canal outside the internal sphincter (Figure 31.1). Posteriorly, the upper fibres of the external sphincter merge with those of puborectalis. Theoretically there are three components to the external anal sphincter complex but these are of little significance surgically. In the lower part of the anal canal there are longitudinal smooth muscle fibres and elastic tissue that extend through the lower fibres of the external sphincter to become attached to the perianal skin. These are extensions of the longitudinal muscle of the rectum.

As the pelvic floor is intimately related with the anal canal, it is also worth considering the anatomy of this structure. The pelvic diaphragm is formed by the levator ani, which arises from the sides of the pelvis and allows the passage of the urethra, vagina and anal canal. The innermost fibres are called puborectalis. This arises from the symphysis pubis and surrounds the vagina or

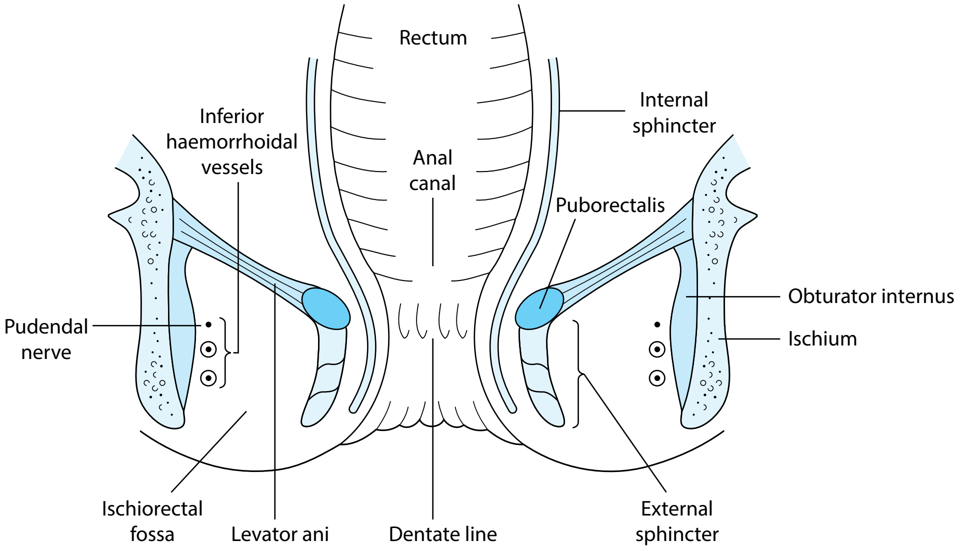


Figure 31.1 Anatomy of the anal canal.

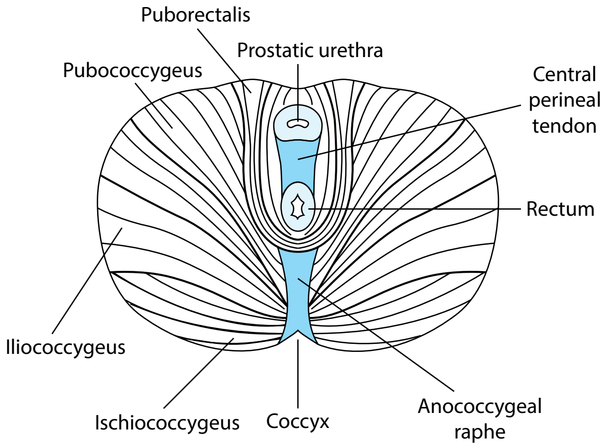


Figure 31.2 Anatomy of the pelvic floor.

prostate and the anorectal junction just above the sphincters and is then inserted into the symphysis pubis on the opposite side. Thus, it creates a sling that is deficient anteriorly. The rest of levator ani is made up of three parts: pubococcygeus, ileococcygeus and ischiococcygeus (Figure 31.2). Pubococcygeus arises from the pubis and is inserted into the coccyx. Ileococcygeus arises from the ileum and is inserted into the tissue between the anal canal and the coccyx (anococcygeal raphe) and into the coccyx itself. Ischiococcygeus arises from the ischial spine and is inserted into the coccyx and lower sacrum. Posteriorly, the pelvic floor is completed by piriformis, which arises on the pelvic surface of the sacrum and is inserted into the tip of the greater trochanter of the femur. On the perineal aspect of the pelvic floor is the central perineal tendon, which is a mass of fibrous tissue lying between the anal canal and the bulb of the penis or vagina. Into this is attached the transverse perineal muscles, which are divided into superficial and deep components. These muscles act as support for the pelvic floor. Deep to these structures lie muscle fibres joining the prostate or vagina to the anal canal that are known variously as levator prostatae, pubourethralis and pubovaginalis.

Nerve supply

Levator ani derives its nerve supply from the third and fourth sacral nerves as they pass from the pelvis through the pelvic floor. These nerves also supply the anal canal and perianal skin. The external anal sphincter complex is supplied by the pudendal nerve, which is derived from S2–4. This nerve leaves the pelvis between piriformis and ileococcygeus and re-enters it through the lesser sciatic foramen. It then runs forwards on the posterior surface of levator ani giving off the inferior haemorrhoidal nerve, the perineal nerve and the dorsal nerve of the penis. The inferior haemorrhoidal nerve supplies the external anal sphincter and perianal skin.

Blood supply

The blood supply to the anal canal comes from the inferior haemorrhoidal artery, which is a branch of the anterior portion of the internal iliac artery. It passes out of the pelvis between piriformis and ileococcygeus and via the greater sciatic foramen and then re-enters by passing over the sacrospinous ligament. It then runs in the ischiorectal fossa to supply the levator ani and sphincter muscles as well as the lower rectum and anal canal. The branches of the vessel that supply the skin of the anal canal have to pass through the internal sphincter muscle.

Venous drainage

The venous drainage very much follows the arterial supply but it is of interest to note that there is a very free communication between superior and inferior haemorrhoidal veins in the submucosal plexus of the anal canal and the rectum. Thus, both the anal canal and the lower rectum can drain directly into the systemic circulation via the internal iliac vein rather than through the portal system.

Lymphatic drainage

The lymphatic drainage of the upper anal canal is via channels around the rectum and then along the inferior mesenteric artery to the preaortic nodes. However, the lower anal canal below the dentate line drains to the perianal plexus and then on to the inguinal lymph nodes.

Physiology

The anal canal is a highly complex mechanism which, under normal situations, allows the individual to control the retention and evacuation of gaseous, liquid and solid rectal matter. There is, however, considerable individual variation and many people will at some point experience incontinence of liquid faeces or flatus. Slight leakage from the anal canal (soiling) is also not infrequent.

When faecal material enters the rectum there are three phases:

- *Accommodation*, in which the rectum slowly expands but both the internal and external sphincters retain their tone.
- *Sampling*, in which the rectal contents come into contact with the sensory lining of the anal canal (transitional zone) after temporary relaxation of the internal sphincter.
- *Defaecation*; although this is under voluntary control to a certain extent, when the volume of rectal contents reaches a critical point the urge to defaecate becomes overpowering and the tone in the external sphincter is inhibited.

The sensory component of incontinence is complex and most individuals are able to distinguish between gas, liquid and solid. Specialized nerve fibres are found within the anal canal and sensory endings in the levator ani complex of muscles have a role in controlling the urge to defaecate. Motor control of continence is exerted by a high-pressure zone in the anal canal (50–100 cmH_2O) produced by the combined tone of the internal and external sphincters. The angle of the anal canal and rectum is about 80° and is maintained by the action of puborectalis muscle. As indicated above, the initial stimulus for defaecation is distension of the rectum. This acts as a spinal reflex via a centre in the lumbosacral region, but cerebral control is exerted over this centre if circumstances are not convenient. Conditioning leads to a degree of cerebral control so that defaecation may take place only once a day even though the rectum contains faeces for much of the time.

During the act of defaecation, particularly in the squatting position, the rectum and anal canal form a straight line owing to straightening out of the anorectal angle. Abdominal pressure is raised and the external sphincter muscle is inhibited, allowing faeces to pass through the anal canal. This may be accompanied by a mass peristaltic action so that the whole distal colon is emptied. In some individuals, however, repeated straining may be necessary to pass several smaller segments of stool.

Investigations

Owing to the complexity of the sphincter mechanism, full investigation of the anal canal and its function requires considerable expertise and specialized equipment. The investigative procedures can be classified under the following headings:

- anal inspection and digital rectal examination
- endoscopy
- ultrasound
- manometry
- rectal sensitivity and capacity measurements
- electrophysiology
- radiology.

Anal inspection and digital rectal (and vaginal) examination

Anal inspection and digital rectal examination can give information about haemorrhoids, presence of anal fissure, perianal fistula, anal or rectal prolapse, rectocele, internal and external sphincter function, puborectal function and anal (and rectal) carcinoma. In women a combined rectovaginal examination is important in assessing a rectocele, anterior rectal cancer or anterior sphincter rupture. This investigation is mandatory for a first assessment.

Endoscopy

The anal canal can be examined endoscopically by means of a short rigid endoscope called, in the UK, a proctoscope (Figure 31.3). This is something of a misnomer, however, and the US term anoscope is more descriptive. This instrument allows visualization of the anal canal and lower rectal mucosa and is particularly useful for diagnosing internal haemorrhoids. It should not be used in the conscious patient when painful anal conditions such as fissure are present.

Ultrasound

Ultrasound is useful for visualizing the anal sphincters. The most widely used probe is a rotating 7 or 10 MHz probe within a water-filled plastic cone (Figure 31.4). This produces a 360° cross-sectional image when inserted into the anal canal. The ultrasound image that is produced images the internal sphincter as a well-defined hypoechoic layer and the external sphincter as

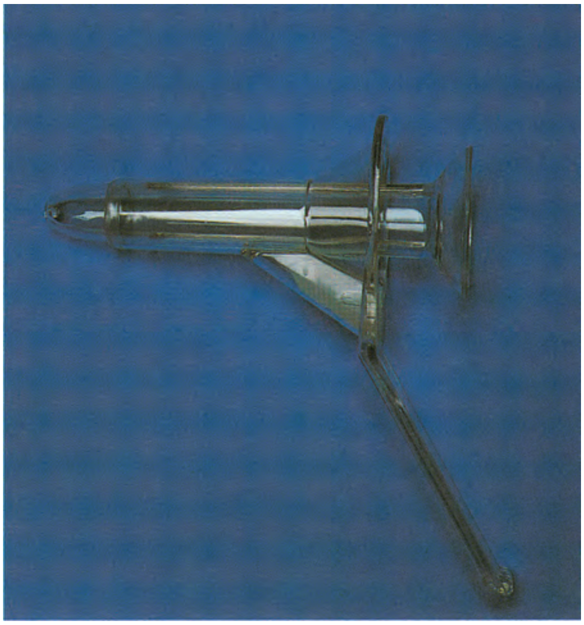


Figure 31.3 Proctoscope.

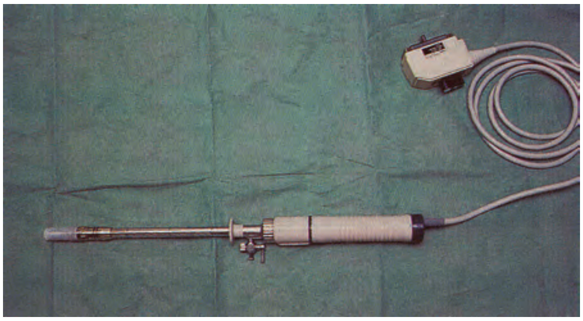


Figure 31.4 Anal ultrasound probe.

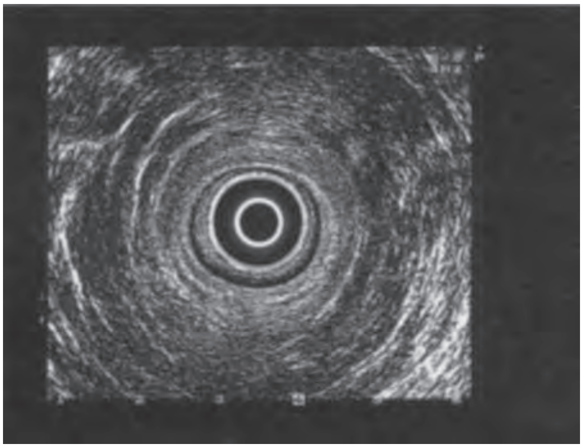


Figure 31.5 Normal ultrasound examination of the anal canal. The internal sphincter is seen as a hypoechoic (black) ring, and the external sphincter as an echogenic (white) ring.

a thick echogenic layer (Figure 31.4). In the lower part of the anal canal the internal sphincter is absent (Figure 31.5) and in the upper part the puborectalis muscle is seen (Figure 31.6). This investigation is very useful in detecting anal sphincter atrophy or anal sphincter disruption. It correlates well with findings during surgery. It is particularly useful in the evaluation of anal fistulas, determining the site and extent of sphincter involvement.

Technology has advanced in this area and the latest probes are oil filled with automated movement and image acquisition (Figure 31.7). This allows three-dimensional reconstruction and greatly increases the ability to later review and interrogate the scan (Figure 31.8).

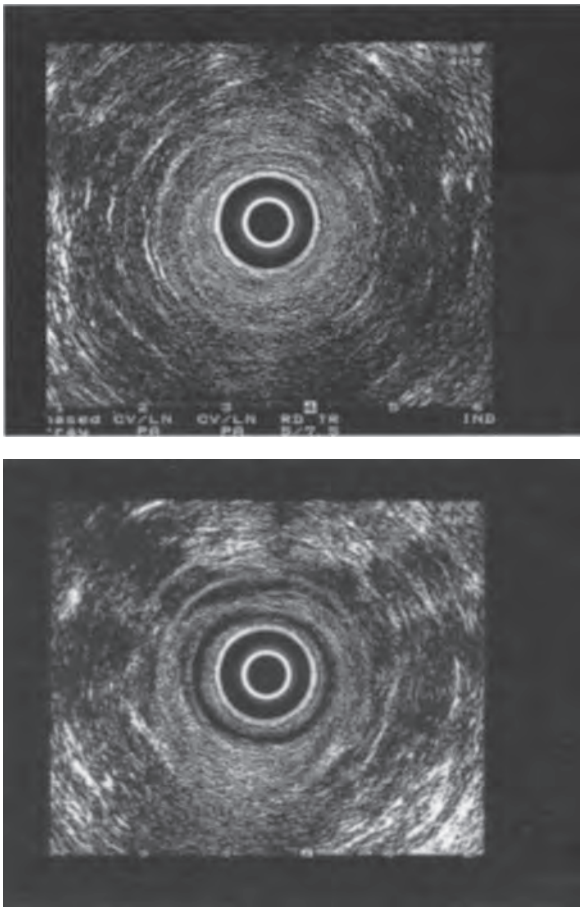


Figure 31.6 (a) Normal ultrasound examination of the lower anal canal; at this level, the internal sphincter is absent. (b) Normal ultrasound examination of the upper anal canal; here, the V-shaped fibres of the puborectalis muscle are seen posteriorly.

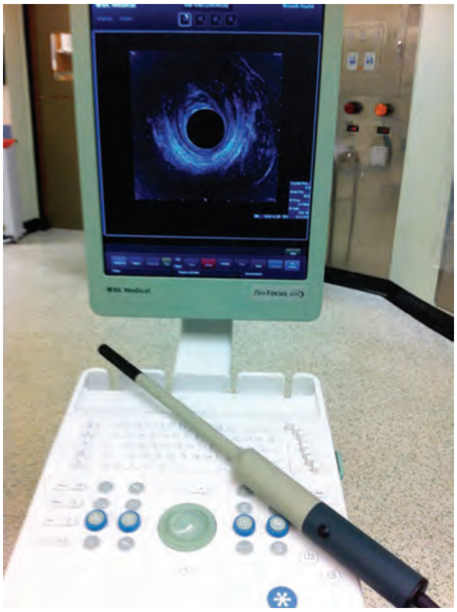


Figure 31.7 Three-dimensional ultrasound probe.

Anorectal manometry

Various devices are available for the measurement of pressures in the anal canal, but the recording device should not be >5 mm in diameter since larger catheters artificially raise anal pressure. The latest technology makes use of strain gauge catheters (Figure 31.9), which contain resistors on a metal diaphragm within a vacuum. Pressure on this diaphragm changes resistance, which is then converted to a pressure measurement. This can be used to produce either a static or an ambulatory measurement. Water-filled systems, however, may retain popularity owing to increased focus on infection control. The resting anal pressure is largely due to the internal sphincter, which is in a state of continuous contraction. This is highly variable, but a pressure of 50–100 cmH_2O is generally regarded as normal. Resting pressure is lower in women than in men and tends to decrease with increasing age. A normal squeeze pressure is in the region

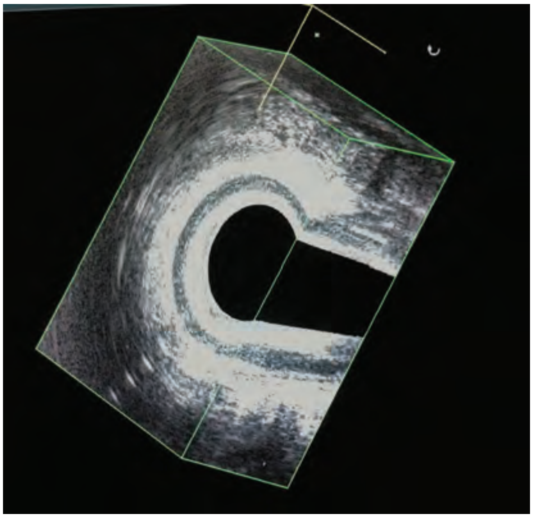


Figure 31.8 Three-dimensional ultrasound images.

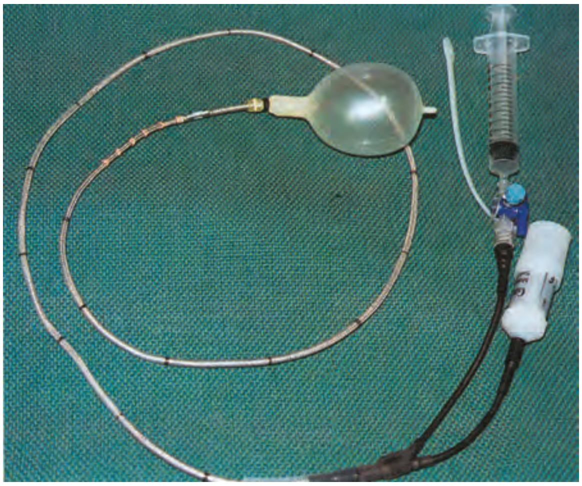


Figure 31.9 Anal manometry catheter.

of 250 cm H_2O for men and 100 cm H_2O for women. In general, patients with faecal incontinence have lower resting and squeeze pressures than continent patients, but there is poor correlation between the pressures and the severity of symptoms. There is also considerable overlap with the so-called normal range. Measures of fatigability may correlate better with incontinent symptoms. The underlying cause for the incontinence cannot be diagnosed on pressure readings alone.

Another useful application of anorectal manometry is testing for the rectoanal inhibitory reflex in patients with defaecation disorders. This is tested by balloon distension of the rectum along with simultaneous recording of the anal pressure. A positive reflex leads to a fall in resting anal pressure of 20%. Although this reflex is absent for a variety of reasons, for example spinal cord injury or anal atresia, a positive result effectively excludes Hirschsprung disease.

Rectal sensitivity and capacity

This investigation also has an important place in disorders of the anal canal. It can be performed by rectal balloon insufflation or a barostat system. The rectal balloon system is easier to use in clinical practice and planning. The advantage of the barostat system is that it can be used to measure rectal compliance, by using a non-elastic bag delivering constant pressures, therefore measuring rectal wall tension.

With the rectal balloon distension measurement, first rectal sensation, first urge sensation and maximal tolerated volume can be measured. In patients with urge incontinence, the first urge sensation is mostly lower than in continent patients. In patients with constipation the rectal sensitivity can be impaired.

Electrophysiology

Faecal or indeed urinary incontinence may be due to spinal or pelvic nerve damage and electrophysiological studies may be useful under these circumstances.

A common problem is incontinence related to child birth, which may be caused by stretch-induced damage to the pudendal nerves. This may be detected by measuring pudendal nerve motor latency. Here a disposable electrode is attached to the gloved index finger (Figure 31.10) and the finger is inserted into the rectum. The pudendal nerves are then palpable as cord-like structures passing around the ischial spines. Latency is defined as the time between the stimulus being delivered to the nerve at the finger tip and the impulse reaching a recording electrode measuring external sphincter contraction at the base of the finger. Normal pudendal motor latency is around 2 ± 0.2 milliseconds. Prolongation beyond the normal range suggests neuropathy and is, for example, associated with a poor outcome following sphincter repair. However, a normal result does not imply lack of neuropathy since only one limited element of nerve function is being tested. Perineal nerve motor latency can also be measured in a similar fashion. Spiral motor latency can be measured by transcutaneous stimulation at the LI and L4 levels with recording of the impulse at the external sphincter. The latency from LI is normally 5.5 milliseconds and from L4 is normally 4.4 milliseconds.

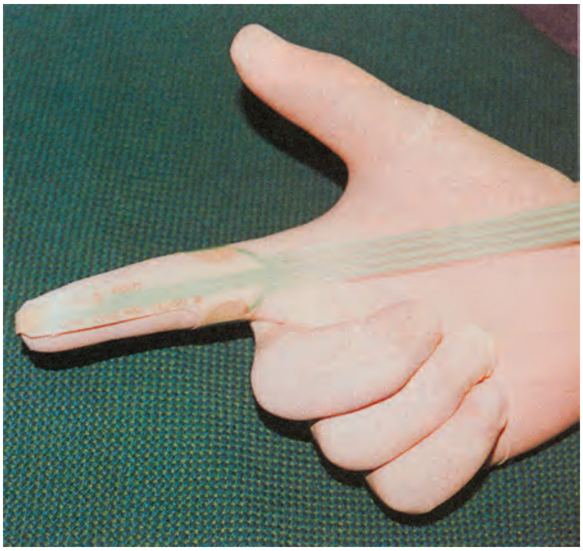


Figure 31.10 Finger-mounted electrode for measuring pudendal nerve motor latency.

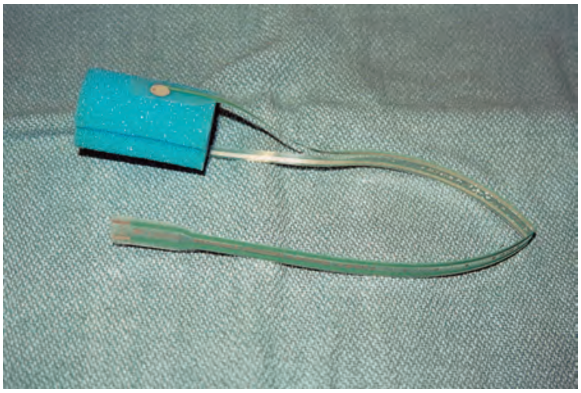


Figure 31.11 Coronal T_2 weighted magnetic resonance image demonstrating a cerebrospinal fluid (CSF) fistula in the left petrous temporal bone with CSF filling the left mastoid air cells (a).

Electromyography involves measuring muscle activity and the most accurate method of testing for denervation is single fibre electromyography. This involves inserting a platinum wire into the muscle. The changes brought about by denervation or reinnervation are quantified by calculating the fibre density, which is the average number of fibres recorded at 20 sites in the muscle. Needle electromyography is not used routinely in clinical practice owing to the discomfort of the test, and difficulty in interpretation. A simple sponge electrode (Figure 31.11) in the anal canal is useful in evaluating abnormal puborectalis contraction during straining (anismus).

Radiology

Defaecating proctography

Anal function is inseparable from rectal function and the most useful form of radiological assessment of this area is defaecating proctography (Figure 31.12). This involves using a thick paste impregnated with barium instilled into the rectum. Also, enteral contrast and vaginal contrast are instilled. The subject then sits on a commode and a video radiograph is taken during evacuation. Under normal circumstances, as straining begins there is a slight concavity of the anterior rectal wall produced by abdominal pressure. The pelvic floor then descends and the anorectal angle widens, the anal canal begins to open, to shorten and to become funnel shaped. A slight degree of rectal wall intussusception can be accepted as normal. This investigation is useful for evaluating rectocele, enterocele, rectorectal intussusception, pelvic floor movement and co-ordination. Thus, it is of value in investigating obstructed defaecation and also in the evaluation of incontinence.

Magnetic resonance imaging

This technique has become the gold standard in the evaluation of complex fistula *in ano*. Integrity and degeneration of the anal sphincter and pelvic floor can also be assessed. Dynamic MRI, giving an insight into the interaction and movement of the pelvic organs during straining and coughing, complements and in some cases may replace defecating proctography. A coil accommodating an upright patient is required for the best results in this type of assessment.

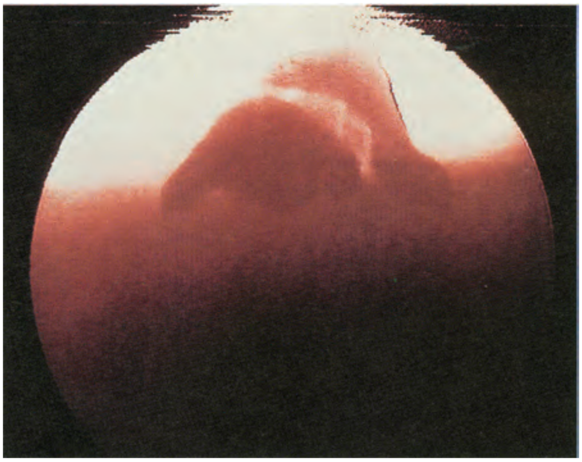


Figure 31.12 Defaecating proctography showing a degree of rectorectal intussusception.

Faecal incontinence

Faecal incontinence is potentially very disabling to patients. It ranges from occasional soiling to regular incontinence of solid faeces. Soiling is minor loss of faecal fluid or mucus confined to staining of underwear or small pads. Faecal incontinence can be either urge or passive incontinence. In the latter, patients are unaware of the need to pass a bowel movement and lose faeces without feeling it, whereas in the former the patient can feel the urge to defaecate but is not able to postpone the bowel movement until reaching a toilet.

Aetiology and pathology

Faecal incontinence can be due to various factors, which are summarized in Box 31.1. It is worth noting that patients with haemorrhoids suffer a mild degree of incontinence, presumably because of a certain amount of sensory impairment in the anal canal owing to loss or displacement of the anal cushions (see Haemorrhoids). These patients do not usually require any intervention, however, and the vast majority of patients presenting with serious incontinence have either an obstetric injury or an iatrogenic injury after anal surgery.

Usually, the woman with incontinence after childbirth gives a history of prolonged labour or a traumatic vaginal delivery usually with forceps. When there has been a severe perineal injury with anal sphincter disruption (third-degree tear) this is usually recognized at the time and immediate repair to the sphincter effected. If this is not recognized, however, the patient will develop severe incontinence soon after delivery. More commonly, the patient will gradually develop incontinence owing to an occult sphincter injury being unmasked by a lax pelvic floor and neuropathic sphincter failure owing to stretching of the pudendal nerves. Thus, the majority of patients with incontinence following single or multiple vaginal deliveries have a combination of nerve damage and direct damage to the pelvic floor.

Incontinence following anal surgery is usually related to laying open of a fistula *in ano* which has transgressed the external sphincter (see Perianal sepsis and fistula *in ano*). Lateral

BOX 31.1 Factors leading to anal incontinence

- Trauma
  - Obstetric injury
  - Surgical injury (after fistula surgery)
  - Accidental trauma (road traffic accident, war casualty)
- Anorectal disease
  - Haemorrhoids
  - Rectal prolapse
  - Anal or low rectal cancer involving the sphincter mechanism
  - Crohn's disease of the anus
- Factors causing diarrhoea
  - Inflammatory bowel disease
  - Infectious colitis
  - Malabsorption
- Congenital disorders
  - Treated imperforate anus
  - Hirschsprung disease
  - Spina bifida
- Neurological disorders
  - Pudendal nerve damage
  - Peripheral neuropathy
  - Stroke
  - Multiple sclerosis
  - Spinal cord damage
- Others
  - Behavioural
  - Faecal impaction
  - Encopresis

Table 31.1 Continence grading scale

Type of incontinence	Never	Rarely	Sometimes	Usually	Always
Solid	0	1	2	3	4
Liquid	0	1	2	3	4
Gas	0	1	2	3	4
Requires pad	0	1	2	3	4
Lifestyle alteration	0	1	2	3	4

Data from Oliveira *et al.* (1996).

sphincterotomy for anal fissure, if done properly, is unlikely to give rise to frank incontinence but may lead to minor problems with control of flatus and mucus (soiling). Occasionally, however, an overenthusiastic sphincterotomy may cause external sphincter damage. Vigorous stretching of the anal canal for fissure or haemorrhoids may also give rise to incontinence owing to combined damage to the internal and external sphincters. Retraction of the sphincter to facilitate endorectal surgery is a risk factor, as is poorly controlled introduction of a circular stapler prior to anastomosis. Injudicious haemorrhoidectomy with excision of the internal anal sphincter is another source of leakage or incontinence.

Clinical assessment and investigation

In the first instance, it is important to establish the severity of the patient's symptoms. Intensive investigations and surgery for minor degrees of incontinence are not usually indicated as they are unlikely to be of value. A scoring system is useful for documenting the severity of incontinence and for estimating the response to treatment (Table 31.1).

A change in bowel habit may unmask an impaired sphincter and this must always be considered. A colonoscopy and biopsies may be required to exclude occult inflammatory conditions. It is also important to establish whether or not there is coexisting urinary incontinence as this suggests a neuropathological cause, and to find out about previous surgery and obstetric history. On examination the patient should be asked to strain in order to assess the presence of a rectal prolapse and the degree of perineal descent. Digital rectal examination will give information regarding resting anal tone and squeeze pressure and it may also be possible to feel a defect particularly in the anterior part of the external sphincter. The history should be a guide to the predominant site of dysfunction. Small quantity passive leakage only suggests an internal anal sphincter-based problem. This should be reflected in a low resting pressure on manometry with preserved squeeze pressure. Ultrasound may reveal deficiency or atrophy of the internal anal sphincter. Pelvic floor exercise is less likely to be of benefit and refractory cases may be considered for anal sphincter bulking injection. Urgency with full blown incontinent episodes implies external sphincter dysfunction. Manometry will generally reflect impaired squeeze pressures and ultrasound may identify a defect in the external anal sphincter. Frequently, the picture is mixed and the contribution of sensorineural changes is being increasingly recognized. Our current ability to accurately assess nerve function is quite limited. Pudendal nerve terminal motor latency and electrosensitivity are very crude assessments.

A trial of properly co-ordinated conservative measures should be the starting point for most patients. If a patient is not responding then a full set of investigations will be required. Ultrasound will document the site and extent of defects in the external or internal sphincter, and reveal occult injuries (Figure 31.13). There is no single objective test that will

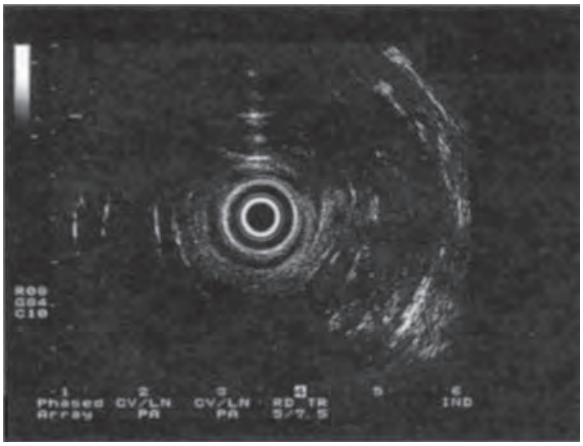


Figure 31.13 Anal ultrasound showing an anterior external sphincter defect. The crescentic echogenic band anteriorly is generated by the examiner's finger in the vagina.

determine the approach in this heterogeneous group of patients. The findings from anorectal manometry, pudendal nerve latency, rectal sensitivity and capacity, and ultrasound of the anal canal require to be critically appraised and correlated with the clinical features in establishing the best way forward for the individual patient. In refractory cases, defecating proctogram or dynamic MRI may suggest a role for an occult prolapsing disorder.

Conservative treatment

Medications

In many patients incontinence is related to loose stool; in such cases, antidiarrhoeal medication may be of value. In patients who are troubled by faecal leakage after defaecation, complete rectal emptying using suppositories (glycerin or bisacodyl) may be helpful.

Pelvic floor exercises and biofeedback

Other conservative approaches include pelvic floor exercises and biofeedback retraining; the combination of both is called pelvic floor rehabilitation. In biofeedback training, the patient is taught how to exercise the external anal sphincter by an anal sensor, and a balloon system is used to train the patient to recognize the stimulus of rectal distension.

Physiotherapists specializing in this area and able to deliver a comprehensive package of measures can effectively treat many incontinent patients without recourse to further intervention. Important areas to address include diet and lifestyle change, manipulation of stool consistency, pelvic floor strengthening, defaecation dynamics and abdominal wall strengthening. Advances in technology have allowed widespread use of electrostimulation and biofeedback retraining with suitable patients able to use the equipment at home.

Anal plug

An anal plug is effectively a modified tampon inserted through the anus. The top opens widely sitting in the rectum and can help some patients by simply blocking off the anal canal. Its use is limited in many patients by discomfort and in others by the device slipping out. Patients with impaired sensation sometimes tolerate the plug well and others find it useful as a back-up strategy in special situations. It is a simple device with little to be lost by a trial of use.

Rectal irrigation

This is a very old technique but regaining popularity. A convenient easy-to-use set of a pump and catheters has been manufactured (Peristeen, Coloplast; Irrimatic, B. Braun) that has made this option more attractive.

Reports are emerging of a possible rectal perforation with the technique, but, for motivated patients willing to persevere and learn to use it for their particular needs, it can be highly successful and by restoring control of their bowel function have a major impact on their quality of life. Overall, the technique is successful in about 60% of patients. It can be particularly helpful in patients with the difficult combination of incontinence and obstructed defaecation. For most incontinent patients retrograde rectal irrigation works well; on occasions antegrade irrigation by means of an appendicocaecostomy, transverse colostomy or sigmoid colostomy is resorted to, as used in refractory constipated patients.

Surgical interventions

Surgery can be divided into approaches to deal with contributory problems such as prolapse, approaches dealing directly with the sphincter complex such as anterior sphincter repair and techniques to augment the sphincter such as gracilis neosphincter.

An alternative is to target the sensorineural system and this approach, sacral nerve modulation, is the one that has advanced the most in recent years. With its lack of any substantial adverse effects many would currently opt for a trial of this approach in almost any patient with significant lifestyle compromise despite full trial of the above detailed conservative measures. The recent emergence of percutaneous tibial nerve stimulation as a potentially less costly means to achieve neuromodulation is of interest and further experience will determine where this will fit into future treatment algorithms.

Surgery for prolapsing disorders

If any doubt persists regarding the role of a prolapsing disorder or the exact nature of the prolapse then an examination under anaesthesia is well worthwhile. Any form of prolapse can potentially lead to incontinence. There are several mechanisms underlying this. Prolapse may repeatedly trigger the sampling reflex, leading to inhibition of internal anal sphincter tone. Many cases of prolapse appear to generate high-pressure waves that overcome sphincter pressure. In full thickness prolapse the sphincter is physically stretched by the prolapse. There is often an accompanying straining disorder leading to perineal descent and pudendal neuropathy.

The technique required to deal with the prolapse must be carefully selected to avoid exacerbating the incontinence problem. When possible, this would be an abdominal technique as perineal procedures will inevitably interfere with rectal compliance and reservoir function (see Chapter 30).

The situation with full thickness prolapse is straightforward in that surgery is clearly necessary and only the technique is a matter of debate. In mucohaemorrhoidal prolapse the situation is less clear. There may be a delicate balance between improving or exacerbating the situation with no test to quantify this.

Some recent interest has surrounded the role of internal intussusception in the aetiology of incontinence as well as in obstructed defaecation. Procedures such as ventral mesh rectopexy offer a more complete restoration of anatomy than posterior rectopexy with better functional results. A great deal of work is required in this expanding area. What is clear is that incontinence and difficulties with evacuation often share common aetiological factors and coexist more frequently than previously appreciated. The interface with gynaecological disorders is expanding. A more sophisticated, wide-ranging, multidisciplinary assessment is going to be required for these patients, but may yield a better outcome for the future.

Haemorrhoids and incontinence

Treatment options in haemorrhoids are dealt with separately. Haemorrhoidal prolapse may be responsible for a degree of relatively small quantity passive faecal leak as well as mucus leakage and irritation. Their exact contribution in a case of incontinence needs to be carefully assessed. The incidence of urgency following stapled anopexy procedures would suggest caution in selecting this approach for a patient with compromise in continence.

Perianal injectable bulking agents

The inner ring or internal anal sphincter muscle keeps the anus closed at all times except when opening the bowel. A treatment has been developed to treat faecal incontinence that involves injecting a substance into or near this muscle to make it bulkier so that the anus closes better. It has been advocated as a simple and safe option.

A silicone biomaterial (PTQ) was shown to provide some advantages and was safer in treating faecal incontinence than carbon-coated beads (Durasphere) in the short term. Similarly, there were short-term benefits from injections delivered under ultrasound guidance compared with digital guidance. Carefully chosen individual patients can certainly benefit from this approach, although evidence-based assessments have not been convincing. Further studies to determine the optimal site, quantity and material for injection are required.

Anterior sphincter repair

In patients with a significant anterior sphincter defect, direct repair can be effective, although initial good results will tend to deteriorate with long-term follow-up. Around 50% of patients have long-term success. The technique involves making an incision anterior to the anal canal and carefully dissecting out the external sphincter down to the levators. Usually there is a fibrous band at the site of the muscular defect (Figure 31.14). This is transected and the cut ends of the sphincter are overlapped and sutured together (Figure 31.15). Many surgeons will supplement this with an anterior levatorplasty, which is carried out by suturing together the two sides of the levator ani muscle. Most surgeons leave the wound at least partially open in view of the high incidence of infection. The patients must have a laxative regime postoperatively to avoid any straining.

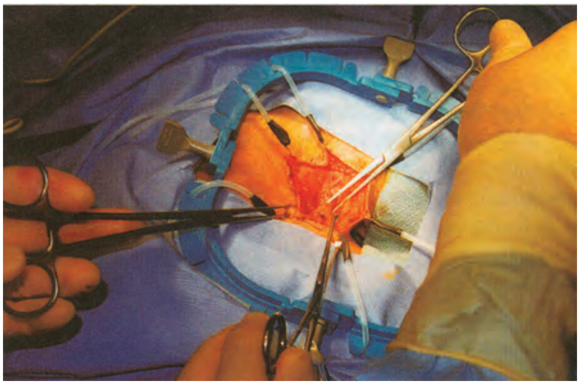


Figure 31.14 Fibrous band at the site of an anterior sphincter defect indicated by the Allis forceps.

Postanal repair

This operation was originally devised in order to restore the normal angle between the anus and the rectum and involves making an incision posterior to the anal margin, dissecting in the intersphincteric plane and plicating the puborectalis muscle. This operation tends to be performed when the sphincter appears to be intact. The early results tend to be good, although the mechanism whereby improvement occurs is controversial. However, long-term follow-up indicates that function deteriorates and the operation has lost favour in recent years.

Total pelvic floor repair

This operation combines a postanal repair with anterior sphincter plication and levatorplasty. It may be more effective than postanal repair, and can be performed in stages or as a single procedure.

Sacral nerve modulation

This is a technique originally developed for urinary incontinence. A low-level electrical stimulus is applied via a percutaneous electrode placed through one of the sacral foramina, preferably the third (Figure 31.16). Although the term sacral nerve stimulation is frequently used, the process by which an effect occurs is unknown. Modulation of reflexes and nerve activity is felt to be the mechanism. The only measurable changes observed

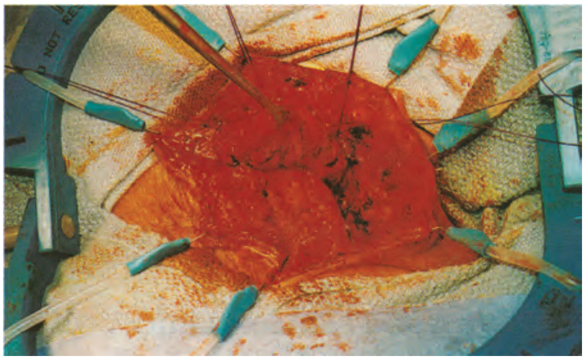


Figure 31.15 Overlapping anterior repair of the external anal sphincter.

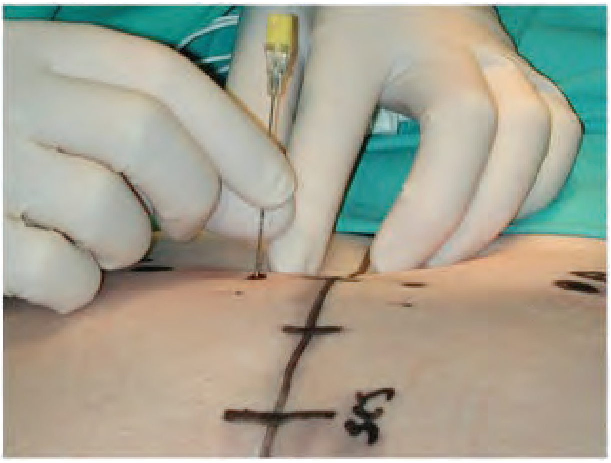


Figure 31.16 Probing sacral foramina for lead placement in sacral nerve neuromodulation.

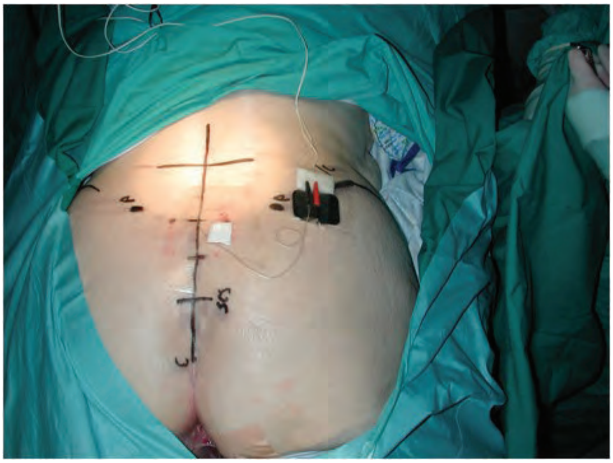


Figure 31.17 Temporary lead in place for test sacral nerve neuromodulation.

are sensory rather than manometric. A test phase (peripheral neural evaluation) is undertaken with an external generator attached to the electrode (Figure 31.17). A sensory response evidenced by a tingling sensation around the anus, or a motor response evidenced by a 'bellows' movement of the perineum, is sought. The patient's clinical response is then evaluated for a period of up to 3 weeks. If an improvement >50% is observed the patient is offered a permanent implant.

A quadripolar electrode is inserted and connected to a pacemaker implanted in the buttock. A success rate of the order of 80% is noted in faecal incontinence, with a very low level of adverse effects. Although the generator is expensive, the test phase allows accurate prediction of which patients will benefit. Sacral nerve modulation is now recommended by the National Institute for Health and Clinical Excellence, which advises on cost-effective evidence-based approaches and provides clinical guidelines in the UK (see Estimating the long-term costs of treatment options in incontinence).

Pudendal nerve modulation

Pudendal nerve modulation is a new technique which may be suitable for patients who have unsuccessful treatment with sacral nerve modulation. Trials are being carried out at the moment.

Percutaneous tibial nerve stimulation

Percutaneous tibial nerve stimulation has come in to use for the treatment of lower urinary tract dysfunction since 2001 and for chronic pelvic pain since 2003. Since 2010 there have been a few studies on treatment in patients with faecal incontinence. In this technique a needle is inserted 3–4 cm proximal to the medial malleolus, between the posterior margin of the tibia and the soleus muscle tendon, and an electrode is placed on the same leg near the arch of the foot. Both are connected to a stimulator. A series of 12 stimulations lasting 30 minutes are applied, separated by no more than 7 days. Success rates of 40–60% have been seen in initial studies.

Dynamic graciloplasty

For many years the gracilis muscle has been used for anal sphincter supplementation (unstimulated graciloplasty), the operation having been originally devised for children with anorectal atresia. In adults, however, straightforward transposition is not associated with particularly good results because the gracilis muscle can only contract voluntarily and is not able to maintain contraction for more than 1 minute. The operation of electrically stimulated gracilis neosphincter, the dynamic graciloplasty, has therefore been developed. The stimulation is provided by two intramuscular electrodes and a stimulator/pacemaker. This operation may be used for patients in whom sphincter repair has failed or is inappropriate, but it may also be used for those who do not have an anal sphincter, e.g. in anorectal agenesis or after abdominoperineal excision of the rectum (Figure 31.18). The results of this operation are difficult to assess objectively but about two-thirds of patients who have undergone the procedure appear to be satisfied with the results achieved. The success rate is lower in patients with anal atresia, most probably because of malformation of neurological, pelvic and sphincter structures in these patients. The technique is surgically demanding with significant morbidity levels. It is best confined to high-volume specialist centres.

Artificial bowel sphincter

An artificial plastic sphincter that can be inflated and deflated by the patient can be inserted around the anal canal (Figure 31.19). The most commonly used device is a modification of the artificial urinary sphincter that is inserted via the perineum. The plastic cuff is connected to a control pump placed in the labia or in the scrotum. This pump is connected to a pressure-regulating balloon

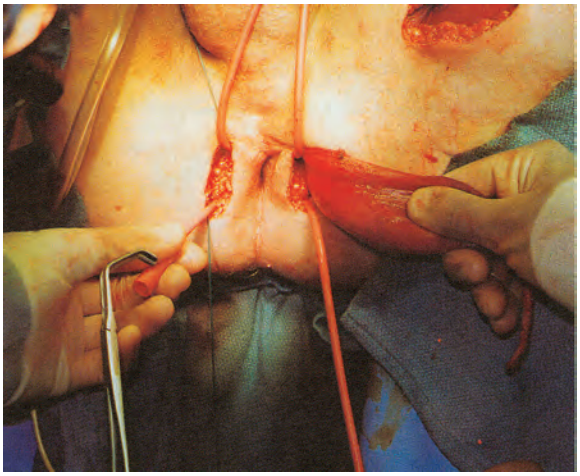


Figure 31.18 Electrically stimulated gracilis neosphincter showing the gracilis muscle mobilized in preparation for plication around the anal canal.

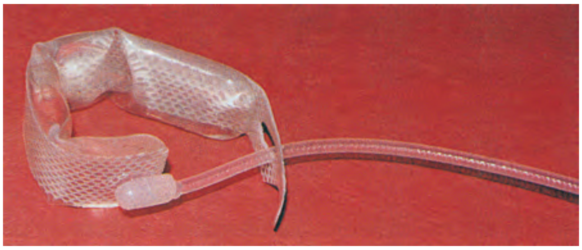


Figure 31.19 Artificial anal sphincter.

placed behind the pubic bone, above the bladder. Success rates of 60% in the long-term can be achieved in terms of continence, although a variety of other problems are experienced by patients. Infection requiring removal of the device occurs in around 20% of cases. Placement of the prosthesis via the abdomen with full mobilization of the rectum may be associated with a lower infection rate, but requires more major surgery.

Although continence is achievable with these sphincter augmentation or encirclement approaches difficulties with evacuation are common and frequently significantly impair the overall quality of life achieved. The sphincter complex has sophisticated dual functionality in terms of maintaining continence and facilitating evacuation.

Estimating the long-term costs of treatment options in incontinence

The 5 year cumulative costs for conservative treatment in faecal incontinence is €3234. The estimated 5 year cumulative cost for colonic irrigation €573. The cost for an anterior anal sphincter repair is €5327. The cost for peripheral neural evaluation and test period are €4053. The 5 year cumulative cost for a sacral nerve modulation implant is €22 150, compared with €33 996 for an end-colostomy, and €31 590 for a DGP. The 5-year cumulative costs of an artificial bowel sphincter is €29 740.

Haemorrhoids

The term haemorrhoids or 'piles' means different things to different people and many patients will use these words to describe a wide variety of anorectal conditions. To the surgeon, however, it refers to abnormalities of the vascular cushions of the anus.

Pathology and aetiology

The anal cushions consist of three spaces filled by arteriovenous communications supported by a fibrous matrix and smooth muscle lying within the anal canal. This allows the anal lining to expand during defaecation but yet to form a complete seal when the anal canal is closed. The arterial supply for these cushions comes from the superior, middle and inferior rectal arteries. Haemorrhoids are thought to result from degeneration of the smooth muscle and fibroelastic tissue that supports the cushions, allowing them to prolapse into the anal canal. However, the underlying reasons for this degeneration are not clear and although constipation and straining at stool have been implicated, the evidence for this is patchy. There is a family history in about 50% of cases and it is therefore possible that a genetic predisposition exists.

Clinical features

The most common symptom is bleeding at defaecation. Commonly this is bright red and follows immediately after defaecation. Typically this is painless but may be quite profuse and frightening for the patient. Other symptoms include perianal swelling, pruritus and minor soiling. Pain from haemorrhoids is associated with complications. Clinically, haemorrhoids can be classified into four groups:

1. internal haemorrhoids presenting with bleeding alone (first degree)
2. haemorrhoids which prolapse on defaecation but reduce spontaneously (second degree)
3. haemorrhoids which prolapse and require manual reduction (third degree)
4. irreducibly prolapsed haemorrhoids (fourth degree).

Thus, on examination the external appearances will depend on the degree of prolapse and the anal canal may in fact look normal. Skin tags around the anal orifice are common and mucosa may be seen to prolapse (Figure 31.20). Digital rectal examination is generally normal. The main diagnostic test is proctoscopy, which gives a good view of the internal anal cushions. It is also essential to examine the rectum with a rigid or flexible sigmoidoscope at least to exclude other lesions.

Investigations

As far as making the diagnosis of haemorrhoids is concerned, investigations other than those mentioned above are unnecessary. However, if there is any doubt about the source of bleeding then a full colonic examination in the form of a flexible sigmoidoscopy and barium enema or a total colonoscopy should be carried out. This would be indicated when there are other symptoms such as change of bowel habit or lower abdominal pain or if the patient is in the high-risk age range for colorectal cancer (i.e. over 50 years of age).

Complications

The complications of haemorrhoids include thrombosis, massive haemorrhage and faecal incontinence.

Thrombosis

When haemorrhoids become irreducible, intravascular thrombosis and oedema may ensue owing to strangulation of the blood supply. This gives rise to severe pain and on examination swollen bluish external haemorrhoids will be seen. Occasionally these may become gangrenous.

Massive bleeding

Very occasionally patients may bleed so profusely from haemorrhoids that they become shocked and require resuscitation

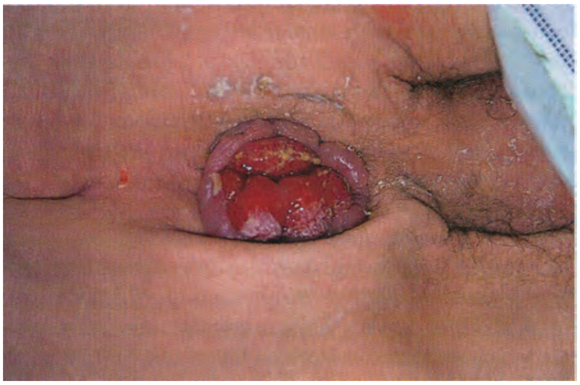


Figure 31.20 Prolapsing haemorrhoids.

with blood transfusion. More commonly, although still relatively unusually, a patient may develop iron-deficiency anaemia from regular bleeding episodes. Attributing iron-deficiency anaemia to haemorrhoids should generally only take place after full investigation to exclude other sources.

Incontinence

Pruritus and minor soiling are relatively common owing to leakage of mucus and liquid faeces from the rectum. This is thought to be due to a poor sealing mechanism owing to displacement of the anal cushions. This may be compounded by a certain amount of sensory impairment in the anal canal. Under most circumstances these symptoms are not particularly disabling but occasionally patients may find them extremely troublesome. If this is the case, the patient should be fully investigated for incontinence as there may be some other underlying cause.

Treatment

The majority of patients with symptomatic haemorrhoids do not need active intervention. Often reassurance after exclusion of serious disease is sufficient. If the patient is finding that constipation or straining is an important feature then bulk laxatives may be of value. Topical ointments may help by providing lubrication but their value is unclear. Diltiazem cream 2% has benefits for reduction of haemarrhoids. It has to be applied twice a day for 6 weeks. If patients have a benefit, they continue for another 6 weeks. Active intervention for haemorrhoids can be divided into two broad areas: (1) outpatient procedures and (2) surgery.

Outpatient procedures

Injection sclerotherapy

For many years injection of sclerosant (most commonly 5% phenol in almond or arachis oil) has been used for the treatment of haemorrhoids. This is injected using a long needle via a proctoscope and 3–5 mL of sclerosant should be injected into the submucosa well above the dentate line at each haemorrhoidal site. The underlying aim is to produce a fibrous reaction within the anal cushion to reduce the degree of prolapse. Care must be taken not to inject too superficially, as this will lead to ulceration, or too deeply, as this will be ineffective. If the injection is too deep it is also possible to damage the prostate or the seminal vesicles and perirectal sepsis has been reported.

Rubber band ligation

An alternative to injection sclerotherapy is rubber band ligation and indeed in randomized trials it has been shown to be more effective. This involves placing tight rubber bands around the prolapsing cushion at least 1.5 cm above the dentate line. There are various devices for achieving this but most are used via a proctoscope. Perhaps the simplest device is a suction tube to which the band is mounted (Figure 31.21). The mucosa is then sucked into the tube and a special triggering device is used to push the band off the end of the tube. More than one band can be inserted at one time although it may be necessary to repeat the procedure. The surgeon should be very careful to apply the band above the dentate line as failure to do this will lead to immediate severe pain. If this happens it is necessary to remove the band by cutting onto it with the tip of a scalpel blade, and this may sometimes necessitate a general anaesthetic. After the procedure patients should be warned to expect some bleeding at between 5 and 10 days when the necrotic cushion separates. They may also expect to have some aching, which may be relieved by warm baths and non-steroidal anti-inflammatory drugs.

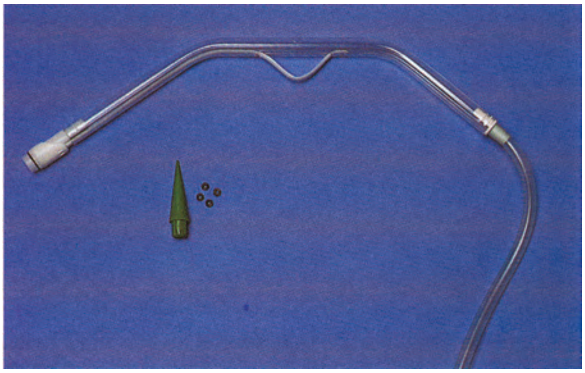


Figure 31.21 Rubber band ligation device for haemorrhoids.

Other outpatient techniques

Bipolar coagulation, infrared photocoagulation, laser photocoagulation and cryotherapy have all been used but none has gained popularity.

The true efficacy of outpatient procedures is not clear. There have been a number of comparative randomized studies that tend to favour rubber band ligation, but unfortunately stratification for severity of disease and the use of no-treatment controls has been lacking. There is little doubt that these procedures have a strong placebo effect and further research is required to establish their precise role in the management of haemorrhoids.

Surgery

In patients who have had failed to benefit from outpatient treatment, the surgical approach becomes necessary. In addition there is a feeling among many colorectal surgeons that patients with severely prolapsing haemorrhoids or in whom bleeding is a major concern should have primary surgical intervention. Currently, there are two widely used surgical approaches: haemorrhoidectomy and stapled anopexy. Haemorrhoidal artery ligation or dearterialization is a more recent approach that is gaining popularity.

Haemorrhoidectomy

There are many different varieties of haemorrhoidectomy but the basic principle is to excise the prolapsing anal cushions while maintaining mucocutaneous continuity between the areas of excision. Each haemorrhoid is grasped close to the mucocutaneous junction in turn and excised in the plane immediately outside the internal sphincter using diathermy. Associated skin tags are included in the excision. If this is done carefully the pedicle will merely consist of a thin strip of mucosa and can be transected directly with diathermy, although some surgeons still prefer to ligate the pedicle. Variations include operating position and whether the mucosa is closed by suture or left open. Lithotomy position is most common in UK practice with prone or left lateral most common in the USA.

Traditionally, after haemorrhoidectomy patients were kept in hospital until their bowels had moved but with careful preparation and community support haemorrhoidectomy can be carried out as a day case. It is even possible to perform the procedure under regional anaesthesia. Steps to diminish pain postoperatively and particularly at the time of defecation include the use of metronidazole, laxative regimens and the use of topical glyceryl trinitrate (GTN) or diltiazem. There is some support for mucosal closure, as in the Ferguson haemorrhoidectomy, compared with the open approach, as in the Milligan–Morgan haemorrhoidectomy, but evidence is not strong in any of these areas.

Stapled anopexy

The term stapled haemorrhoidectomy is inaccurate since the intention is not to excise the prolapsed haemorrhoidal cushions but relocate and fix them; the term stapled anopexy is more appropriate, although less popularly applied. The procedure has attracted a great deal of interest in the last few years. The principle of this operation is to carry out excision of a circumferential strip of mucosa above the dentate line and to simultaneously close the defect. This pulls the mucosa and therefore the anal cushions back up into their normal position, thus restoring the anatomy of the anal canal. This is done using a specially designed proctoscope and circular end-to-end anastomosing stapler.

A purse string is inserted in the rectal mucosa 3 cm above the dentate line, the stapler is inserted and the purse string tightened around the centre rod (Figure 31.22). The stapler is then fired, simultaneously excising the mucosa and stapling the two cut ends together (Figure 31.23). After excising a cam 'doughnut' a circumferential staple line should be left 1 cm above the dentate line. If the stapling is performed too proximally it will be ineffective in elevating the anal cushions, and if performed too distally will risk interference with the sphincter complex. Immediately after the procedure the staple line must be inspected for bleeding points, which can be oversewn. In theory, confining the surgery to the less sensate area above the mucocutaneous junction should be less painful and randomized controlled trials have consistently shown an advantage to the stapled technique in terms of pain. Haemorrhoidectomy has an unfortunate reputation for pain among the general public with many postponing or avoiding intervention. Although certainly not painless, stapled haemorrhoidectomy does appear to increase patient acceptability and may help facilitate day-case surgery. In expert hands the technique appears at least as effective as excisional haemorrhoidectomy but there may be an increased risk of recurrent prolapse over time. It has not been possible to demonstrate a functional advantage despite a more restorative approach. Although large numbers of cases have now been performed uneventfully, care is required as perforation and complete closure of the rectum have occurred (see also Haemorrhoids and incontinence).

Haemorrhoid artery ligation/dearterialization

The latest approach in the quest for an effective but painless treatment for haemorrhoids involves identifying and ligating the arterial inflow to the haemorrhoidal pedicles. This is achieved by using specially designed proctoscopes incorporating a Doppler device and facilitating guided haemorrhoidal artery suture and ligation. The latest variation allows expansion of the technique to treat haemorrhoidal prolapse by suture fixation (rectoanal repair).

The technique is disseminating rapidly with several supportive publications. It appears to be well tolerated and have minimal adverse consequences so far.

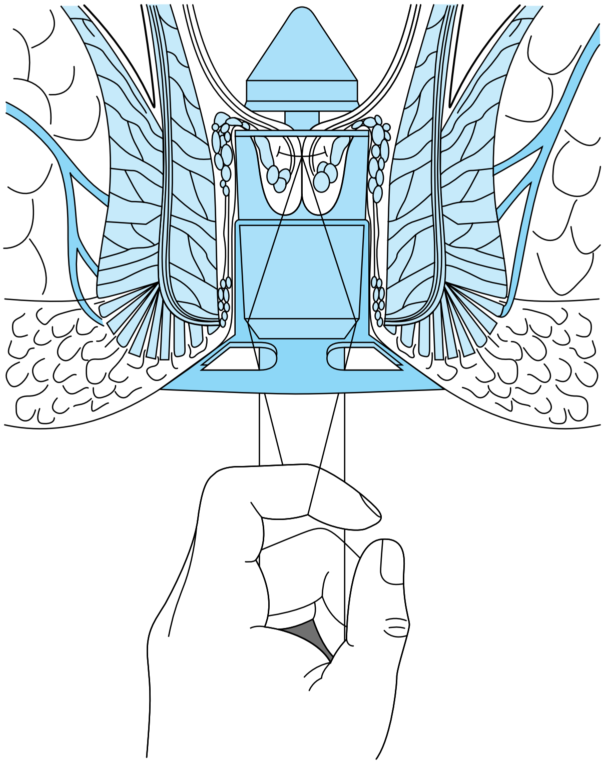
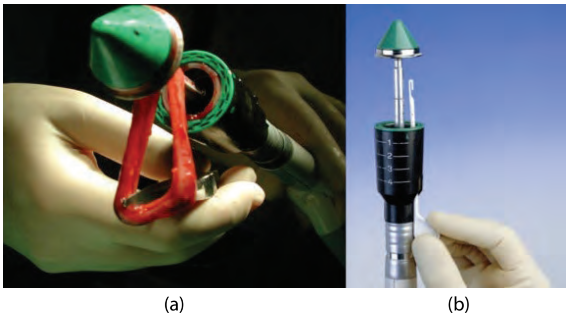


Figure 31.22 Stapled anopexy. Prolapsing rectal mucosa is drawn into a circular stapling gun. Simultaneous excision and stapling draws the anal canal back into an anatomical position.



(a) (b)

Figure 31.23 Stapled anopexy device with excised doughnut (a). Hook for retrieving purse-string sutures into stapler housing (b).

Treating complications of haemorrhoids

The patient with strangulated thrombosed haemorrhoids usually requires hospitalization for adequate analgesia and bed rest. Cold compresses applied directly to the haemorrhoids are also beneficial. Surgeons are divided as to whether or not early haemorrhoidectomy should be carried out in these patients and often an individual decision has to be made on the basis of the severity and duration of the symptoms. Those against intervention argue that following an episode of strangulation haemorrhoid symptoms often resolve spontaneously. Occasional cases of portal pyaemia have been reported.

A case of prolapsed thrombosed haemorrhoids need to be distinguished from that of a thrombosed external haemorrhoid. This is another misnomer as the latter condition is simply a haematoma forming in relation to the external haemorrhoidal plexus and quite separate from the anal cushions. Incision and evacuation of the clot will rapidly relieve symptoms.

Massive bleeding usually requires haemorrhoidectomy and should be distinguished from the occasional case of variceal bleeding secondary to portosystemic shunting at the anorectal junction. Minor degrees of incontinence usually respond well to carefully performed haemorrhoid surgery.

Perianal sepsis and fistula *in ano*

Perianal sepsis and fistula formation may be associated with a number of disease processes including Crohn's disease, malignancy, tuberculosis, pilonidal sinus and trauma. In the majority of patients, however, the condition is idiopathic.

Pathology and aetiology

Central to current thinking regarding perianal sepsis and fistula formation are the anal glands. These glands are situated in the intersphincteric space and open into the anal canal at the dentate line via a duct that transverses the internal sphincter. The function of these glands is not clear but as they secrete mucin they may have a lubricant function. It is thought these glands may become infected if the duct becomes blocked and, when this occurs, pus accumulates within the gland. The pus may then track superiorly, inferiorly, laterally or circumferentially.

Most commonly, the pus will pass downwards in the intersphincteric plane to form a perianal abscess. It may also find its way through the external sphincter into the ischiorectal fossa and thus form an ischiorectal abscess. More rarely the pus may tract up in the intersphincteric plane and form an intersphincteric abscess or discharge down into the ischiorectal fossa through the levator ani muscles (Figure 31.24). When a perianal or ischiorectal abscess discharges through the skin, either spontaneously or as a result of surgical intervention, it may resolve completely. However, if the duct between the gland and the dentate line remains patent and becomes infected, the patient may then be left with a fistulous communication between the dentate line and the skin. According to the mode of spread the fistulous tracts can be classified in the following way (Figure 31.25):

- intersphincteric
- trans-sphincteric
- suprasphincteric
- extrasphincteric.

Intersphincteric fistulas make up about 50% of all fistulas and usually consist of a straightforward tract between the dentate line and the skin incorporating part of the internal sphincter. However, some of these can have a high intersphincteric extension and even a high opening into the rectum.

Trans-sphincteric fistulas account for about 30% and consist of a tract passing through the external sphincter. This may be low or high and occasionally may be associated with a blind high tract in the ischiorectal fossa which may even penetrate the levator ani muscles.

Suprasphincteric fistulas run above the puborectalis muscle and then descend down through the levator ani muscles into the ischiorectal fossa.

Extrasphincteric fistulas bypass the sphincter complex completely and extend from the lower rectum through the levator ani muscles and into the ischiorectal fossa.

This classification is important as it has implications for treatment.

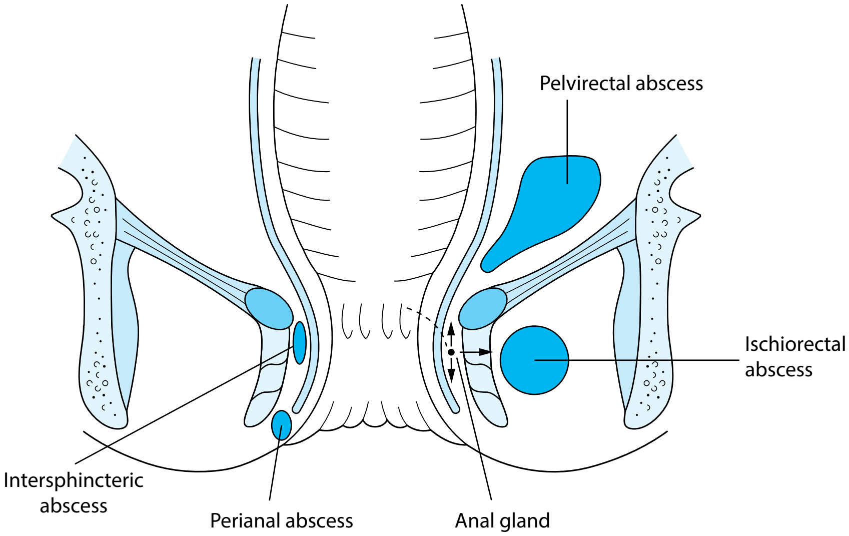


Figure 31.24 Anorectal abscesses.

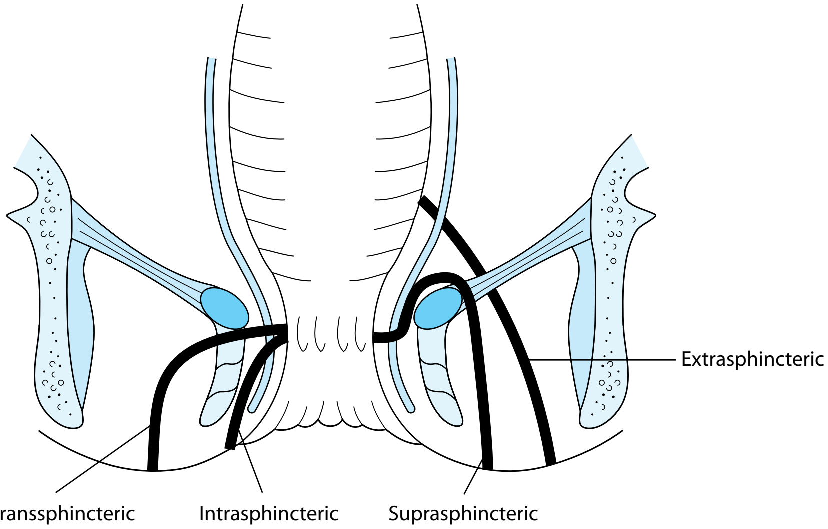


Figure 31.25 Anorectal fistulas.

Clinical features

The patient with a perianal sepsis usually presents with severe perianal pain, fever and malaise. The abscess may discharge spontaneously, but because of the severity of symptoms patients often present at hospital before this has occurred. On examination a perianal abscess is usually very obvious as a tender red swelling at the anal margin. The ischiorectal abscess, however, may be quite deep seated and more difficult to detect on clinical examination, although pressure on the ischiorectal fossa will usually give rise to severe pain. In an advanced ischiorectal abscess a large area of tender induration will be seen. Occasionally, a patient may present with an intersphincteric abscess that has not pointed at the anal canal and this can be very difficult to detect clinically. Thus, in the patient with a severe acute anal or perianal pain an examination under anaesthetic should be carried out. If there is an intersphincteric abscess this will be felt as induration through the rectal wall.

A fistula may present after discharge or incision of an abscess, usually after all the inflammation and induration has settled down in about 50% of all abscesses. Alternatively, there may be no history of abscess formation and the fistula may appear to arise *de novo*. The main symptoms are of discharge and pruritus, although occasionally a patient may notice the passage of flatus through a fistula track. There may also be a history of recurring episodes of pain relieved by discharge from the fistula.

On examination a punctate opening (or openings) can be seen, usually close to the anal verge, although sometimes a few centimetres away. Pus or serosanguinous discharge may been seen exuding from the opening particularly if digital pressure is applied over the fistula. Careful firm palpation between the opening and the anal canal will often reveal the fistulous track as a subcutaneous 'cord'. Goodsall's rule (Figure 31.26) is useful when estimating the course of a fistula; this states that, if a fistulous opening is posterior to an imaginary line drawn transversely through the middle of the anus, the track will curve round so that it opens into the dentate line in the posterior midline, whereas, if the opening is anterior, the track will be radial (i.e. it follows a straight line from the opening to the dentate line). The main exception to this rule is the anterior opening, which is ≥3 cm from the anal verge, as this may be 'horseshoeing' round from the posterior midline.

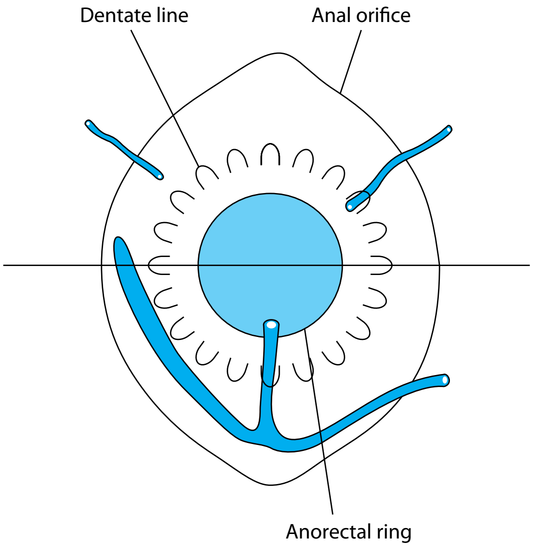


Figure 31.26 Goodsall's rule related to anal fistulas suggests that fistulas with an external opening anterior to a line drawn horizontally through the anal canal progress forwards in a radial fashion, whereas fistulas with an external opening posterior to the horizontal line curve backwards and ultimately have an internal opening in the midline posteriorly. It is possible for fistulas to extend laterally on both sides, leading to the characteristic horseshoe fistulas.

Diagnosis

The diagnosis of an abscess is usually made clinically, and its exact position relative to the sphincter complex is made at operation. Likewise, the course of a fistula is usually established by examining the patient under general anaesthesia using specially designed fistula probes. The most widely used probes are those designed by Lockhart-Mummery; these are available in a variety of configurations, and the most useful are slightly curved (Figure 31.27). The flat handle allows for precise manoeuvring of the probe, and the groove is useful for laying the fistula open (see below).

Under normal circumstances it is relatively easy to pass the probe from the external opening along the tract to the internal opening at the dentate line (Figure 31.28). In doing this it is important not to use too much force, as it is possible to create a false tract in so doing. If it proves to be impossible to find the internal opening using this technique a valuable manoeuvre is to instil a very small amount of hydrogen peroxide into the external opening by means of a fine cannula. With an Eisenhammer retractor in place within the anal canal it is then usually possible to see bubbles appearing at the site of the internal opening. This will guide further probing. In the process of probing a fistula it is important to look for extensions to the main tract. Indeed, if the probe cannot easily pass along the tract it is probably falling into a blind extension bypassing the main tract.

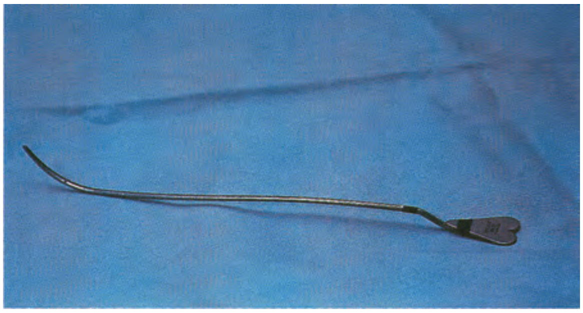


Figure 31.27 Lockhart-Mummery fistula probe.

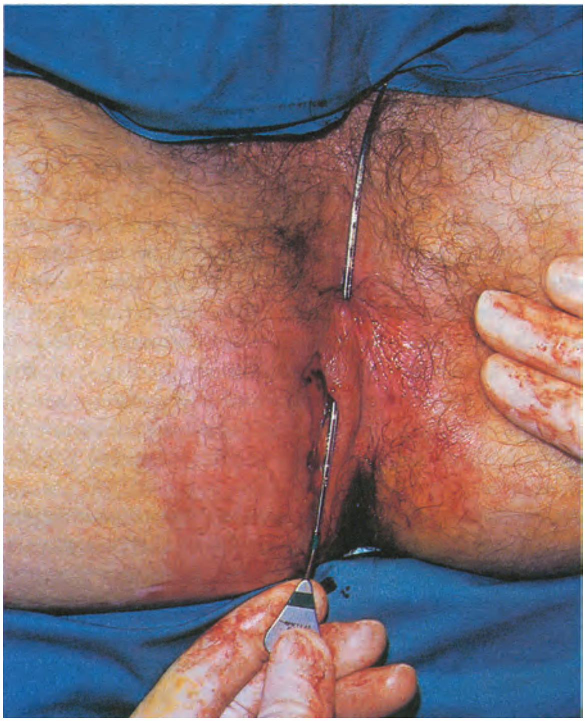


Figure 31.28 Probe in a fistula *in ano*.

Although examination under anaesthetic with probing is the most useful diagnostic approach, some forms of imaging may also be of value. Fistulography has been used in the past. However, its accuracy is sparse. Currently, fistulography has been superseded by endosonography and MRI (Figure 31.29). It appears to be particularly useful in complex fistulas with secondary extensions and abscesses and, although it is perhaps no more accurate than an examination under anaesthetic by an experienced coloproctologist, a high-quality MRI of the sphincter complex may be useful as a 'road map' to guide the surgeon. Transanal ultrasound is undertaken using a 10 mHz mechanically rotated transducer to give a 360º axial image.

Treatment

The initial treatment of a perianal or ischiorectal abscess which is pointing onto the skin is fairly straightforward. The abscess cavity should be incised and drained and an opening in the skin made that is large enough to allow continued drainage of the abscess. The cavity should not be packed tightly as this makes it uncomfortable for the patient and actually impedes drainage of the abscess cavity. A large cavity may require dressing changes in hospital until it can be managed by a community nurse. Alternatively, it is possible to put a counterincision in a large abscess and leave a drain. The patient now has to irrigate the wound in the shower twice a day but can be managed as an outpatient. The patient should always be followed up in the outpatient department to look for the development of a fistula when the cavity itself has healed. One of the complications of incision and drainage of a perianal or ischiorectal abscess is the iatrogenic production of a high fistula by overenthusiastic curetting of the abscess cavity or injudicious probing for a fistula. This must be avoided at all costs.

The best treatment for fistula *in ano* is undoubtedly a fistulotomy, laying open of the entire tract, curetting out of

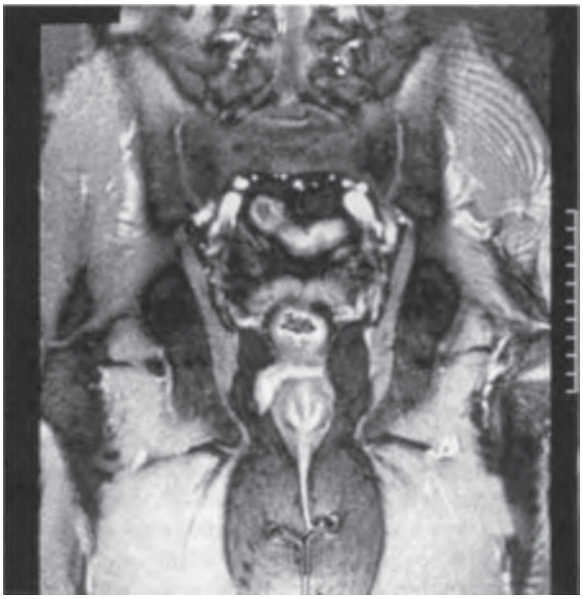


Figure 31.29 MRI scan of a fistula *in ano*. The fistula is seen as a white track curving around the anal canal. (Courtesy of Dr D. Shephard, Dundee, UK.)

the granulation tissue and leaving the wound to granulate. This is done most easily by cutting down onto the groove on the concave aspect of the fistula probe, and then using a small Volkmann's spoon to curette. Some surgeons will marsupialize the tract by suturing the divided wound edge to the edges of the fibrous tract and this is said to result in faster healing. A fistulotomy is generally safe with a simple intersphincteric fistula as the only muscle that will be divided will be the lower part of the internal sphincter. If, however, the fistula is trans-sphincteric then a fistulotomy creates a risk of incontinence. A fistulotomy of a suprasphincteric or extrasphincteric fistula will inevitably result in complete incontinence.

In the case of a low trans-sphincteric fistula that only involves a small part of the lower external sphincter, a fistulotomy is relatively safe. However, at the time of the initial examination it is important to assess exactly how much of the external sphincter is involved. It is therefore safer to establish drainage of the fistula by inserting a seton through the fistula tract. This can simply be a length of suture material or a vascular sling loosely tied. The seton controls the situation by ensuring ongoing drainage through the tract and preventing recurrent abscess formation. When the patient is awake after the procedure it is then easier to assess the anal sphincter and, if the surgeon is confident that only a small part of the external sphincter is involved, the patient can go back for laying open of the fistula.

If, however, a significant part of the external sphincter muscle is involved or if the fistula proves to be suprasphincteric, then a more conservative approach must be taken. The most widely used approach is to ensure complete drainage of all the sepsis and to leave the seton in place for several weeks. When all the sepsis and inflammation has resolved the seton can then be removed and this will result in healing in about 50% of cases. If this fails, the seton must be reinserted and a careful discussion with the patient must take place. If fistulotomy will leave some external sphincter muscle intact then this is a feasible option as long as the patient is aware that a certain degree of incontinence may result. The situation is always more difficult when some pre-existing sphincter compromise is at play and for women in general, in whom a shorter sphincter and the potential adverse effect of childbirth on the continent mechanism require to be considered.

When a seton is left in place for a longer time, it may work its way towards the surface. In this way the high fistula becomes a low fistula. The seton gradually cuts its way through the fistula tract leaving fibrosis behind it. Although the sphincter is divided more slowly than in fistulotomy the effect on continence is not very much different with some degree of incontinence in about 60% of cases.

An alternative procedure for the trans-sphincteric fistula is fistulectomy and advancement flap repair. Here the fistula tract is excised or cored out by following it up from the external orifice and dissecting through the external sphincter muscle. A flap of mucosa and internal sphincter is then raised above the internal opening and sutured down over it. This is successful in some cases, but the overall rate of success is difficult to estimate. Repeating the procedure can increase the success rate, but repeated interventions with scarring and fibrosis will in themselves impair anal canal function. A degree of laxity in the mucosa is helpful in raising a flap and may account for the lower success rates in males and in posterior versus anterior fistulas.

With a high trans-sphincteric fistula or a suprasphincteric fistula incontinence is inevitable with fistulotomy, and one approach is to give the patient a temporary colostomy, lay open the fistula and then carry out a sphincter repair after healing has taken place. Alternatively the patient may opt to live with a long-term loose seton in place. This approach is the safest in a variety of situations, including multiple fistulas and Crohn's disease.

A variety of approaches have been taken to try and induce healing of anal fistulas without resorting to fistulotomy or fistulectomy. In general, the aim has been to fill the tract with biological material in the hope that it will be slowly replaced by the body's own tissue, leaving a healed fistula tract. Injection of fibrin glue was used quite extensively. Initial results seemed promising but long-term results have been poor with high recurrence rates. The liquid consistency of fibrin glue is possibly not ideal for the purpose of closing anorectal fistulas, because the glue is easily extruded from the fistula tract by increased intraluminal pressure. A plug fabricated from porcine collagen, which is claimed to stimulate tissue remodelling leading to closure of the fistulous tract, is currently being used. The tract is cleaned out, the plug is pulled into position and then sutured in place with the mucosa closed over the top (Figure 31.30). A multicentre randomized controlled trial is now taking place in the UK comparing this plug with a cutting seton approach (Fistula In Ano Trial – FIAT).

The extrasphincteric complex fistula, which is usually secondary to Crohn's disease or trauma, is particularly difficult to deal with and, particularly in Crohn's disease, it may be better to avoid surgical interference. In such patients, a long-term seton may be the answer and some patients may require a defunctioning colostomy. After resolution of the sepsis, laying open of the fistula and subsequent sphincter repair may be possible, but the ultimate results can be less than ideal.

Current interest surrounds the LIFT (ligation of intersphincteric fistula tract) procedure. This involves dissection upwards in the intersphincteric space to the level of the fistula which is then ligated and divided. The external fistula is curetted. The technique is attractive as it has minimal impact

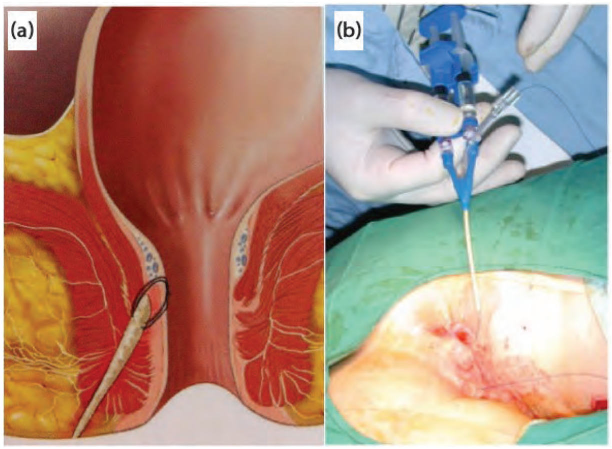


Figure 31.30 Fistula plug (a) and fibrin glue (b).

on the sphincter muscle. A variation is to insert a small sheet of biomaterial between the divided ends of sphincter (BioLIFT). Early results are encouraging but wider experience and long-term follow-up are required to fully assess its role.

Pilonidal sinus and abscess

Pilonidal sinus is a condition that occurs mostly in the natal cleft of young males. It is characterized by multiple subcutaneous sinuses and abscess cavities containing hair.

Aetiology and pathology

There is very little evidence for a congenital origin in this condition and current consensus favours an acquired mechanism. It is thought that frictional forces generated in the depths of the natal cleft tend to drive hairs subcutaneously, where they generate a foreign body reaction. Secondary infection may compound the problem, leading to abscess formation.

Clinical features

Pilonidal sinus may be asymptomatic and only present on routine inspection. However, if the sinuses become infected the patient may have pain and discharge in the natal cleft. On examination there are a variable number of pits seen in the upper natal cleft (Figure 31.31) and, in active inflammatory disease, these may be seen to be exuding pus. When an abscess forms there will be a red tender swelling just to one side of the natal cleft. Although the appearances are typical, occasionally a pilonidal sinus may be mistaken for a fistula *in ano* and vice versa. The diagnosis is made on clinical grounds and specific investigations are not generally required. Sometimes a granuloma can be seen on excised specimens.

Treatment

If asymptomatic pits are found then no treatment is required. If a patient presents with a pilonidal abscess then this should be incised and drained, and the patient kept under review until the wound has healed. Thereafter, a decision can be made regarding excision. The most difficult situation to treat is the patient with chronic discharging sinuses, and this requires a fairly radical surgical approach.

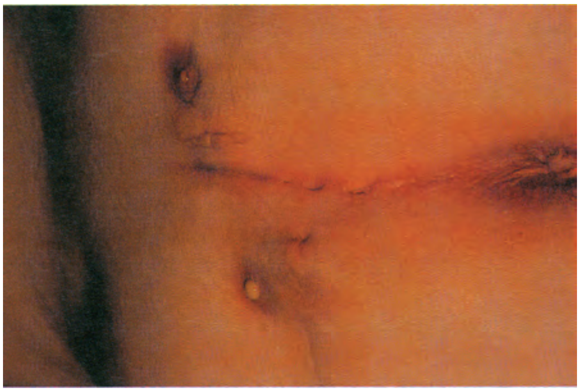


Figure 31.31 Pilonidal sinus. Midline pits in the natal cleft with distant fistulous openings. (Courtesy of Mr R.A.B. Wood, Dundee, UK.)

The traditional approach has been excision of the sinus-bearing area down to the deep fascia and removal of all sinuses and nests of hair. Surgical opinion is then divided as to whether this should be left open to granulate or whether primary closure should be attempted. If the latter approach is employed almost 50% of the wounds will break down and healing by secondary intention will be necessary. In a patient who has pits that are some distance from the main disease in the natal cleft an elliptical excision of the main area may be carried out in association with 'coring out' of the long extensions. These need to be left open to heal by secondary intention. The long-term recurrence rate after excision of pilonidal sinus is in the region of 20%.

Many attempts have been made to design a definitive procedure for this condition. The plethora of approaches reflects the lack of a single procedure combining a proportionate level of complexity with reproducible efficacy.

Interest has focused on relocating the suture line away from the midline and obliterating the midline cleft. Procedures such as those described by Bascom, Karydakis and Limberg share this aim using a variety of approaches. These techniques do have better results, but can be extensive.

Hidradenitis suppurativa

Apocrine glands are found in certain zones including the axillae, the inguinoscrotal and perianal regions and the breasts. These glands develop from hair follicles and discharge a thick secretion into the follicle or onto the adjacent skin. As they are not active until puberty, inflammation of these glands (hidradenitis) does not usually appear until the third decade. Clinically, the affected area becomes indurated then forms sinuses with the discharge of small amounts of pus. Although the axillae are the most predominantly affected areas the perianal region is involved in about 30%.

Conservative treatment is the first approach. The patients are advised to wear cotton underwear and avoid shaving with razorblades. Betadine scrubs can be helpful. During active periods the hidradenitis may require treatment with antibiotics. Sometimes long-term rotating courses are required. When severe, treatment requires excision of the affected skin and subcutaneous tissue down to the deep fascia, and, when extensive, split skin grafting is necessary to provide cover. If the perianal region is extensively affected it may be necessary to carry out a temporary defunctioning colostomy to allow excision and skin grafting. There is a strong tendency to recurrence of the disease.

The development of squamous cell carcinoma has been described in patients with long-standing perianal hidradenitis suppurativa.

Perianal Crohn's disease

Anywhere between 30% and 70% of patients with Crohn's disease may have some involvement of the anal canal. This is more common in patients with colonic and particularly rectal Crohn's disease, but it may occur in isolation or in association with Crohn's disease elsewhere in the gastrointestinal tract.

BOX 31.2 Classification of Crohn perianal lesions (after Hughes)

- Primary lesions
  - Anal fissure
  - Oedematous piles
  - Ulceration
- Secondary lesions
  - Skin tags
  - Anal stricture
  - Perianal abscess and fistula
  - Anovaginal and rectovaginal fistula
  - Carcinoma
- Incidental lesions
  - Piles
  - Skin tags
  - Perianal sepsis
  - Hidradenitis suppurativa

The perianal lesions are variable and have been classified by Hughes (Box 31.2); according to this classification the primary lesions consist of anal fissure, ulcerated oedematous haemorrhoids and deep ulceration of the anal canal or lower rectum. These can lead to the secondary lesions of oedematous skin tags, anal or rectal stricturing, perianal abscess and fistula formation, and rectovaginal or anovaginal fistula. It must also be remembered that Crohn's disease is potentially a premalignant condition and carcinoma of the rectum or anal canal may complicate this disease.

The severity of perianal Crohn's disease is highly variable and may cause only minor irritation. On the other hand, in a patient with multiple fistulas ('watering can' anus) the severity of the symptoms may necessitate proctectomy. The management of perianal Crohn's disease is conservative where at all possible. The disease is characterized by poor healing following surgical intervention. Metronidazole is particularly useful but can require prolonged courses, which risk the onset of neuropathy. Ciprofloxacin is also associated with good results, but use recently has been tempered by the incidence of *Clostridium difficile* infection.

Surgical intervention is required for eradication of sepsis and this involves draining abscesses and inserting setons into fistulas. Frequently, the fistulas are high and may be extrasphincteric so that laying open of a Crohn's fistula is usually contraindicated. Occasionally, with severe disease, proctectomy is necessary, but this should only be carried out after sepsis has been controlled; otherwise, an infected perineal sinus is likely to result.

New biological therapies and advances in MRI mean that the assessment and management of perianal Crohn's disease should now require a multidisciplinary approach involving a surgeon gastroenterologist and radiologist. There have been some encouraging results in the treatment of Crohn's anal fistulas with antitumour necrosis factor (anti-TNF) antibody therapy, but it is likely that sustained therapy is necessary to maintain improvement. In addition, caution is required as increased abscess rates have been observed and radiological assessments demonstrate persisting fistula appearances even when clinical healing has been achieved. An examination under anaesthesia is usually required to eradicate any obvious ongoing sepsis prior to commencing anti-TNF therapy. The exact management of a seton suture during therapy remains to be established. Early removal will risk abscess formation whereas delayed removal may prevent healing. Further studies on the length of therapy, dosing interval, second antibody and other factors are required to establish what the place of these therapies in long-term healing of Crohn's anal fistulas might be.

Anovaginal fistulas are particularly distressing and will usually require intervention. Good control of Crohn's disease combined with advancement flap repair gives the best chance of success but recurrence is common. Bringing in healthy tissue for interposition such as gracilis muscle may be necessary.

Anal fissure

An anal fissure is a linear ulcer that occurs in the anal canal just distal to the dentate line. This may be caused by Crohn's disease or trauma but most commonly it is a primary condition. It affects both men and women and the highest incidence is in the third and fourth decades of life. It may occur soon after pregnancy and vaginal delivery.

Aetiology and pathology

The initiating factors in anal fissure are unclear, although minor anal trauma caused by passage of a constipated stool has been suggested. The main underlying pathology, however, appears to be a high resting anal pressure caused by increased internal sphincter tone. The blood supply to the anal canal has to pass through the internal sphincter and therefore spasm of this muscle reduces the blood flow and the oxygen tension in the skin of the anal canal. Interestingly, the fissures tend to occur at the watershed of the blood supply, i.e. the anterior and posterior midline in women and the posterior midline in men.

Clinical features

The typical clinical features are of pain on defecation associated with bright red bleeding. This may be associated with pruritus ani and discharge of mucus. On examination there is usually a skin tag overlying the fissure and the fissure itself can be seen by everting the anal canal using lateral traction (Figure 31.32). This will reveal a sharply defined ulcer and it may be possible to see the lower fibres of the internal sphincter at its base. Digital rectal examination or proctoscopy should not be attempted in the conscious patient as this will cause considerable discomfort.

Diagnosis

In terms of making a diagnosis this is done clinically, but it is important to exclude other conditions such as Crohn's disease or malignancy. This can be established by examination under anaesthesia and biopsy when appropriate.

Treatment

The underlying principle of treating anal fissures is to reduce the internal anal sphincter tone. In patients with minimal

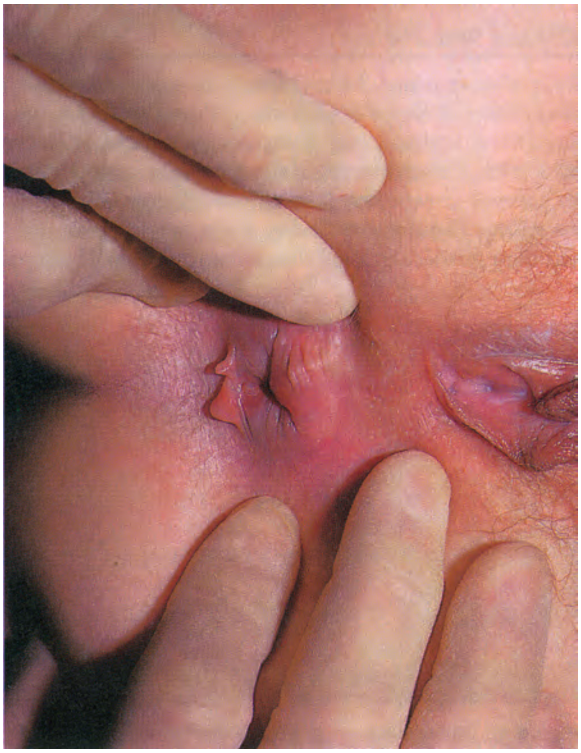


Figure 31.32 Anal fissure.

symptoms this may be achieved by topical application of a local anaesthetic and bulk laxatives. In patients with more severe symptoms, however, the use of 0.2% GT) cream applied two or three times a day can produce healing of fissures in about 50% of cases. Topical diltiazem has a similar effect but may be better tolerated. These agents probably act by both relaxing internal anal sphincter and improving the blood flow owing to their vasoactive properties. The healing takes about 6–12 weeks. For the patient unable or unwilling to use topical application oral nifedipine can have a similar effect.

If these approaches fail then a trial of temporary paralysis of the internal anal sphincter by botulinum toxin type A injection is worth trying and has the advantage that the effects will wear off without leaving any permanent sphincter compromise. The dosage and site of injection (intersphincteric versus intramuscular) is debated as is the exact mode of action of botulinum toxin in this situation. Few adverse effects have been reported and this approach may reasonably be positioned between topical treatments and surgical intervention. If this does not work the first time, it can be repeated after 6 weeks.

If medical treatment and botulinum toxin injection fail then a surgical approach becomes necessary. Historically, forced anal dilatation (Lord procedure) was performed and, although effective, this was associated with an unacceptable level of incontinence. The surgical treatment of choice is now a lateral sphincterotomy, which involves dividing the internal sphincter at one point on the lateral wall of the anal canal up to the level of the dentate line. This is achieved by inserting an anal retractor so that the internal sphincter is gently stretched and easily palpable. A small incision is then made on the lateral aspect of the anal canal just below the internal sphincter. Using scissors the intersphincteric plane is developed, as is the plane between the anal skin and the internal sphincter. The scissors are then used to divide the sphincter and bleeding is controlled by firm finger pressure. Lateral sphincterotomy is said to be successful in about 95% of cases, but patients should be warned that it can be associated with minor degrees of incontinence to flatus or mucus. Adjustments to the surgical technique to diminish the incidence of incontinence include never extending the sphincter incision above the top of the fissure or beyond the dentate line.

An alternative approach is required in refractory cases and when there is pre-existing sphincter compromise. These are frequently cases in which sphincter pressure is low as opposed to the ultra-high pressures usually noted in anal fissure. These cases merit full investigation by anorectal physiology and anal ultrasound. Ultrasound may reveal a fully intact internal anal sphincter in cases of failed lateral sphincterotomy and suggest that the procedure was inadequate and may be repeated. Otherwise, a mucosal advancement flap can be used with a fissurectomy or the 'house' advancement flap, which aims to cover the exposed internal anal sphincter fibres by advancing skin from the anal verge upwards to cover the excised fissure bed.

Pruritus ani

Pruritus ani or perianal itching is a common condition. There are a large number of anorectal and dermatological conditions that may give rise to this symptom, but in the vast majority it is idiopathic. The primary cause is probably minor anal leakage as the result of internal sphincter dysfunction. This sets up an irritation that is exacerbated by brisk cleansing or scratching. This may then introduce fungal infection that intensifies the pruritus. Another important cause is use of wet, perfumed toilet paper, giving rise to chemical irritation or allergic reaction.

When taking a history it is particularly important to establish whether the family includes small children as this would predispose to *Enterobius* infestation (threadworm). On examination the perineal skin should be closely inspected to look for signs of dermatological disease. In idiopathic pruritus there is typically perianal excoriation and ichthyosis. Digital examination of the rectum should be carried out in order to assess the anal tone and proctoscopy to look for the presence of haemorrhoids. Perianal lesions or areas of abnormal skin should be biopsied and, when appropriate, skin scrapings should be examined for fungal infestation.

If *Enterobius* infestation is suspected, the Sellotape test should be carried out. This involves placing adhesive tape over the anus and then transferring it to a glass microscope slide. Histological examination will subsequently reveal the presence of ova that have been deposited on the perianal skin.

The treatment of pruritus ani depends on any underlying cause. Prolapsing haemorrhoids, anal polyps and skin tags should be treated surgically. *Enterobius* infestation can be treated using mebendazole. If there is a primary skin condition the patient should be referred to a dermatologist. In the case of idiopathic pruritus advice should be given in order to minimize anal leakage and damage to the perianal skin. This involves dietary modification to reduce excessive flatulence or loose stool. The patient should also be advised to keep as clean and dry as possible and to avoid scratching. In case of some leakage, rectoanal irrigation can be performed by a small irrigation bottle. When the pruritus is intense and particularly if a fungal infection is suspected then a cream containing steroid and an antifungal agent (e.g. clotrimazole) can be used in the short term to break the cycle of itching and scratching.

In extreme cases of pruritus ani, ablation of the sensory nerves by subdermal injection of methylene blue has been used.

Perianal warts

Perianal warts or anal condylomata are caused by infection with human papillomavirus (HPV) and is a sexually transmitted disease. It is found in between 40% and 70% of males who have sex with males. Subtypes of the virus that are implicated in the development of these lesions are 6, 11, 16 and 18. The most common is subtype 6, but subtypes 16 and 18 are important in that they are associated with a high incidence of dysplasia and malignant transformation. In women, about 80% will have associated warts on the external genitalia, vagina or cervix and about 20% of men will have warts on the penis.

Clinically, patients tend to present with perianal itching and bleeding, and on examination warts are seen in the perianal region with varying degrees of confluence (Figure 31.33). In patients with AIDS, the warts tend to form a dense confluent sheet.

There are various methods of treatment, and, for relatively mild disease with isolated warts, 25% podophyllin in either liquid paraffin or tincture of benzoin can be applied. Intervening normal skin should be avoided as podophyllin is very irritant and may give rise to skin necrosis. In general, although podophyllin treatment is effective the results are not as good as surgical excision, which is the preferred method.

Excision involves operating on the patient in the prone jack-knife or lithotomy position and a 1:200 000 solution of epinephrine (adrenaline) is injected subcutaneously. By causing swelling in the perianal skin, this separates the warts and they can be excised individually with scissors. The rate of recurrence after complete excision of all perianal warts is in the region of 10%. Unfortunately, this approach can be difficult when the warts are extremely confluent and it may be necessary to resort to electrocoagulation or cryotherapy. Both of these result in quite severe tissue damage and anal stenosis is a possible complication. Recently, an autologous vaccine prepared from the patient's own warts has been described and this appears to be highly effective not only in eradicating the warts but also in reducing the rate of recurrence. Interferon has also been used to good effect.

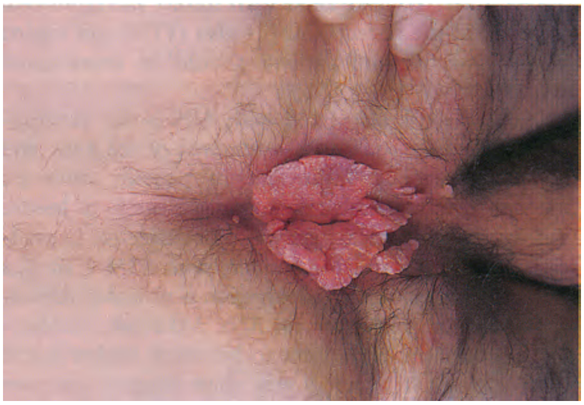


Figure 31.33 Perianal warts.

Anal intraepithelial neoplasia

As is the case with perianal warts, anal intraepithelial neoplasia (AIN) is caused by HPV and appears to be sexually transmitted. HPV types 6, 11, 16, 18, 31 and 33 can all be implicated. On histological examination epithelial dysplasia is seen and this is graded as 1–3 according to the epithelial depth of dysplasia. Thus, in grade 1 only the upper third is affected; in grade 2 the upper two-thirds; and in grade 3 the whole thickness of the epithelium is dysplastic. Grade 3 AIN is also classified as carcinoma *in situ*.

Clinically, patients present with perianal discomfort, itching and bleeding and on examination the perianal epithelium is thickened and may have a whitish appearance. This may be exaggerated by the application of acetic acid, which is useful in targeting areas for biopsy. High-resolution anoscopy is a new promising instrument to detect lesions early.

In women, it is important to appreciate that AIN may be associated with similar intraepithelial neoplasia of the cervix, vulva and vagina and affected patients should be investigated appropriately.

The main risk associated with AIN is the development of invasive squamous carcinoma of the anus, and the treatment and follow-up of patients must be directed towards early detection of such a lesion. Unfortunately, surgical excision is not effective in eradicating the virus and for patients with AIN 1 or 2, a simple policy of regular observation is probably regarded as adequate. However, for AIN 3 the risk of developing carcinoma is thought to be much higher and if there is a small affected area then surgical excision would seem to be appropriate. However, when there is a large circumferential area of AIN 3, treatment is much more difficult. Some have recommended complete excision of the affected area with split skin grafting. However, this requires the formation of defunctioning colostomy and is associated with a high risk of anal stenosis. In addition, it does not effectively eradicate the virus. Imiquimod, which acts as an immune response modifier, is reported to have a good effect on AIN, but further studies need to be conducted. Currently, therefore, it is recommended that such patients be followed up 6 monthly with careful inspection and biopsy with intervention only if invasive malignancy develops.

These patients should be treated by a multidisciplinary team, including a gynaecologist and a dermatologist.

Anal cancer

Anal cancer accounts for only about 5% of all carcinomas of the large bowel and it is nearly always a squamous carcinoma. Rarely, adenocarcinoma may be seen but this is usually a very low rectal cancer involving the anal canal. Carcinomas near the dentate line are mostly mixed squamous cell carcinomas and adenocarcinomas (adenosquamous carcinomas). Another very rare condition is malignant melanoma of the anal canal. This section will, however, deal exclusively with squamous carcinoma. Histologically anal cancer is termed epidermoid and this may be subclassified as squamous cell, basaloid (or cloacogenic) or mucoepidermoid. This classification does not appear to have any prognostic significance, because there is no difference between the squamous cell carcinoma and the adenosquamous carcinoma in regard to treatment, response to treatment and prognosis. Local spread of an anal cancer tends to occur proximally and radially into the rectum and anal sphincters. In advanced cases it may also involve the vagina. Lymphatic spread is to the perirectal and inguinal lymph nodes. Distant metastases occur most frequently in liver, lung and bones but this is usually associated with advanced local disease.

Aetiology

It is now clear that the majority of anal cancers are associated with HPV infection, although not all patients will have a clear-cut history of perianal warts or AIN. The relationship seems clearer for cancers from within the canal as opposed to from the perianal skin. The carcinogenic factors involved are not clear, but immunosuppression or HIV infection appear to increase the incidence of anal cancer.

Clinical features

Patients tend to present with anal pain and bleeding and some will notice a mass in the anal canal. On examination the typical appearance is of a malignant ulcer at the anal margin or within the anal canal. It is also important to examine the inguinal region, although when inguinal lymph nodes are enlarged only about 50% will have metastatic spread on histological examination. Delay in presentation occurs when symptoms are attributed to haemorrhoids. Atypical features and disproportionate pain should always arouse suspicion.

Investigation

Clearly the most important investigation is biopsy of the anal lesion, but examination under anaesthetic is extremely important in evaluating the extent of disease and its resectability with primary surgery. MRI of the pelvis with gadolinium and endorectal ultrasound are useful in assessment of the primary cancer. A CT scan of the chest, abdomen and pelvis is required for full staging of the disease.

Classification

The treatment of anal carcinoma is based on the location of the tumour and the tumour (T)–node (N)–metastasis (M) classification.

Location

A squamous cell carcinoma of the skin more than 2 cm from the dentate line, without involvement of anal canal, needs to be treated as a squamous carcinoma of the perineal skin. A squamous cell carcinoma of the skin within 2 cm of the dentate line with involvement of anal canal needs to be treated as a squamous carcinoma of the anal canal.

Treatment

This is as follows:

- Perianal skin
  - Tis: local excision; if margins are not clear then adjuvant radiotherapy.
  - T1N0M0: local excision with a 1 cm margin; when the margin is too small, there should be adjuvant radiotherapy to the primary tumour field.
  - T2 (≤4 cm) N0M0: radiotherapy to the primary tumour and inguinal lymph nodes.
  - T2 (>4 cm) to T4N0M0: chemoradiotherapy for the primary tumour and inguinal lymph nodes.
  - T2–4N1–3M0: chemoradiotherapy to the primary tumour and lymph nodes.
  - T1–4N0–3M1: individualized approach.
- Anal canal
  - Tis: local excision; when margins are not clear use adjuvant radiotherapy.
  - T1–2 (≤4 cm) N0M0: radiotherapy to the primary tumour.
  - T2 (> 4 cm) to T4N0M0: chemoradiotherapy for the primary tumour and pelvic and inguinal lymph nodes.
  - T1–4N1–3M0: chemoradiotherapy to the primary tumour and lymph nodes.
  - T1–4N0–3M1: individualized approach.

Surgery retains an important role in the treatment of anal cancer. About 35% of patients undergoing combined radiochemotherapy will either experience recurrence or fail to fully respond to the treatment. In this case, radical surgery in the form of abdominoperineal excision of the rectum is necessary. In addition, some patients may have such a debilitating tumour that defunctioning colostomy is necessary while they are having the radiochemotherapy. Although chemoradiation is effective and may avoid a stoma, patients require full information on the long-term effects, especially in relation to bladder and sexual function. Younger patients may opt for a primary surgical approach. When there is residual tumour after primary surgical resection chemoradiotherapy is indicated.

Follow-up

A programme of regular follow-up is required to optimize early detection of residual tumour after primary treatment, local recurrence and inguinal lymph node metastases. Targets of later follow-up include treatment side effects such as radionecrosis and secondary tumours after chemoradiotherapy.

GUIDE TO FURTHER READING

Alexander-Williams J. Pruritus ani. *BMJ* 1983;**287**:159–160.

Cirocchi R, Farinella E, La Mura F, *et al*. Fibrin glue in the treatment of anal fistula: a systematic review. *Ann Surg Innov Res* 2009;**3**:12.

D'Haens GR, Panaccione R, Higgins PD, *et al*. The London Position Statement of the World Congress of Gastroenterology on Biological Therapy for IBD with the European Crohn's and Colitis Organization: when to start, when to stop, which drug to choose, and how to predict response? *Am J Gastroenterol* 2011;**106**:199–212.

Faucheron, J-L, Poncet G, Voirin D, *et al*. Doppler-guided hemorrhoidal artery ligation and rectoanal repair (HAL-RAR) for the treatment of grade IV hemorrhoids: long-term results in 100 consecutive patients. *Dis Colon Rectum* 2011;**54**:226–223.

Harmston C, Jones OM, Cunningham C, Lindsey I. The relationship between internal rectal prolapse and internal anal sphincter function. *Colorectal Dis* 2011;**13**:791–795.

Henry MM, Swash M (eds). *Coloproctology and the Pelvic Floor.* London, UK: Butterworths, 1985.

Mowatt G, Glazener CMA, Jarrett M. Sacral nerve stimulation for faecal incontinence and constipation in adults. *Cochrane Database Syst Rev* 2009;(3):CD004464.

Mundy L, Merlin TL, Maddern GL, Hiller JE. Systematic review of safety and effectiveness of an artificial bowel sphincter for faecal incontinence. *Br J Surg* 2004;**91**:665–672.

National Institute for Health and Clinical Excellence. *Faecal Incontinence*. Clinical Guideline 49. London, UK: NICE, 2007.

Oliveira L, Pfeifer J, Wexner SD. Physiological and clinical outcome of anterior sphincteroplasty. *Br J Surg* 1996;**83**:502–505.

Parks AG, Gordon PN, Hardcastle JD. A classification of fistula-in-ano. *Br J Surg* 1976;**63**:1–12.

Phillips RKS, Lunniss PJ. *Anal Fistula. Surgical Evaluation and Management*. London, UK: Chapman & Hall, 1996.

Shao W-J, Li G-C H, Zhang Z H-K, *et al*. Systematic review and meta-analysis of randomized controlled trials comparing stapled haemorrhoidopexy with conventional haemorrhoidectomy. *Br J Surg* 2008,**95**:147–160.

Surgery

Significant changes in surgical training have occurred over the past decade, with the inevitable progress towards specialization in the various surgical specialties. The current edition of *Essential Surgical Practice: Higher Surgical Training in General Surgery* reflects these latest developments with a new structure that accurately mirrors current surgical training. This new edition excludes surgical specialties that are distinct from general surgery, allowing an increased focus on topics of direct relevance to trainees in higher surgical training.

Over 31 chapters, the book provides detailed information on those specialties of which the general surgeon is required to have a working knowledge, and highlights core information for revision and quick reference. Topics include:

- Surgical biology and pathology
- Surgical craft approaches and technologies
- Perioperative care
- Cardiovascular, pulmonary and renal pathophysiology
- Surgical infections
- Interventional radiology
- Trauma and head injuries
- Disorders of the thyroid and adrenal glands
- Disorders of the stomach, liver, spleen, colon and pancreas
- Bariatric surgery

This book can be used with confidence by those enrolled in local courses and is also consistent with the scope and level of information required for international postgraduate examinations, such as the Joint Surgical Colleges' Fellowship Examination. To enhance the book's utility, the print edition includes complimentary access to a VitalSource ebook online and offline on your PC or Mac, iPhone®/iPod Touch®/iPad®, Kindle Fire or Android™ device.

Professor Alfred Cuschieri **FRSE, MD, DSc, ChM, FRCS, FACS (Hon), LLD (Hon), FMedSci, FiBiol,** Senior Scientist, Institute for Medical Science and Technology, University of Dundee, Dundee, UK; Professor of Surgery, Scuola Superiore Sant'Anna, Pisa, Italy

Professor George B Hanna **MD, PhD, FRCS,** Head of the Division of Surgery, Imperial College London, UK

WITH VITALSOURCE® EBOOK

- download the ebook to your computer or access it anywhere with an internet browser
- search the full text and add your own notes and highlights
- link through from references to PubMed

K17645

CRC Press
Taylor & Francis Group
an **informa** business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

ISBN: 978-1-4441-3760-6
90000
9 781444 137606